Mississippi State University
Scholars Junction

Theses and Dissertations | Theses and Dissertations

1-1-2016

Dimensionality Reduction of Hyperspectral Imagery Using Random Projections

Vineetha Menon

Follow this and additional works at: https://scholarsjunction.msstate.edu/td

Recommended Citation
Menon, Vineetha, "Dimensionality Reduction of Hyperspectral Imagery Using Random Projections" (2016). *Theses and Dissertations*. 1511.
https://scholarsjunction.msstate.edu/td/1511

This Dissertation - Open Access is brought to you for free and open access by the Theses and Dissertations at Scholars Junction. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.

Dimensionality reduction of hyperspectral imagery using random projections

By

Vineetha Menon

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Electrical and Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2016

Copyright by

Vineetha Menon

2016

Dimensionality reduction of hyperspectral imagery using random projections

By

Vineetha Menon

Approved:

James E. Fowler
(Major Professor/Graduate Coordinator)

Qian (Jenny) Du
(Co-Major Professor)

Robert J. Moorhead
(Committee Member)

Nicolas H. Younan
(Committee Member)

Jason M. Keith
Dean
Bagley College of Engineering

Name: Vineetha Menon

Date of Degree: December 9, 2016

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Dr. James E. Fowler

Title of Study: Dimensionality reduction of hyperspectral imagery using random projections

Pages of Study: 80

Candidate for Degree of Doctor of Philosophy

Hyperspectral imagery is often associated with high storage and transmission costs. Dimensionality reduction aims to reduce the time and space complexity of hyperspectral imagery by projecting data into a low-dimensional space such that all the important information in the data is preserved. Dimensionality-reduction methods based on transforms are widely used and give a data-dependent representation that is unfortunately costly to compute. Recently, there has been a growing interest in data-independent representations for dimensionality reduction; of particular prominence are random projections which are attractive due to their computational efficiency and simplicity of implementation. This dissertation concentrates on exploring the realm of computationally fast and efficient random projections by considering projections based on a random Hadamard matrix. These Hadamard-based projections are offered as an alternative to more widely used random projections based on dense Gaussian matrices. Such Hadamard matrices are then coupled with a fast singular value decomposition in order to implement a two-stage dimensionality

reduction that marries the computational benefits of the data-independent random projection to the structure-capturing capability of the data-dependent singular value transform. Finally, random projections are applied in conjunction with nonnegative least squares to provide a computationally lightweight methodology for the well-known spectral-unmixing problem. Overall, it is seen that random projections offer a computationally efficient framework for dimensionality reduction that permits hyperspectral-analysis tasks such as unmixing and classification to be conducted in a lower-dimensional space without sacrificing analysis performance while reducing computational costs significantly.

DEDICATION

To my Family and Friends for their relentless faith and support

ACKNOWLEDGEMENTS

BE the Change that you want the world to SEE !

– Vineetha Menon

I set out on my humble journey in the US strongly based on my set of ideals and beliefs that I would eventually carve out a niche to establish myself in this big world. I often remind myself of one of the famous Zen teachings with profound insight regarding the importance of lucidity and open-mindedness in our lives—"Be not the rigid rocks that obstruct the flow of water in a river, but be the languid water that can slide through creeks and with perseverance can erode the rock itself." Perseverance has always been the key for my success. Credited to my numerous interactions with a diverse community in MSU, I have been able to develop a greater perspective of things in life, while still staying true to my core beliefs. I guess, this is what people deem as *experience* or *wisdom* (in lackluster terms—*growing old!*), and it would truly be incomplete without the remarkable people in my life. My Ph.D. expedition has been an exciting roller-coaster ride so far!

I extend my thanks and gratitude to my major adviser, Dr. James E. Fowler, for all your advice, patience through phases of productivity and drought alike, and most importantly encouraging and backing me in all situations. You have greatly influenced my professional ideals. Much thanks and appreciation for my co-major adviser, Dr. Qian Du, for all the support, promptness, insightful discussions and assisting me in selecting my research

topic. A special note of thanks for my committee members, Dr. Nicholas H. Younan, for your relentless support and encouragement, and Dr. Robert J. Moorhead for your valuable insights and comments. And thanks to all the ECE faculty and staff, for your exemplary service. I'm extremely proud to be a part of MSU-ECE-Bulldawg family!

My deepest gratitude and appreciation for my family for all the love, relentless faith and prayers; my dearest Grandpa, Dr. Raymond S. Winton, and his family, for all the care, love, wonderful times, and most importantly putting up with my headstrong nature! You have influenced my life beyond words can express. You are and will be my cherished extended family for life. A special note of thanks to my dearest friends Dr. Shantia—for your wise counsel, friendship and mentorship, Sheida and Fataneh (& family) for all your hugs and love—you brought me back to life! You are all treasured family and are a part of me forever!

Special thanks to my wonderful research group: Dr. Sungkwang and Dr. Nayeon—thank you for being my angels and restoring my faith in humanity during tough times, Dr. Eric and Laura—for all the discussions, movie nights and my first baseball game!, Dr. Nam and Trang—for your precious friendship and our fun-filled panoramic multi-"you" experiments, Dr. Wei and Dr. Chen for your valuable discussions, friendship and joyful times!, Dr. Azam—for your witty and happy company, Dr. Zhen—for our shopping trips and volunteering to be my guinea pig-first passenger in my car!, Dr. Miao Zhang—for your dear friendship and medicines when I was sick, Dr. Deng, Dr. Wen and Liu—for all the happy times and cheers to our movie trips! I'm truly fortunate to know you all and heartiest thanks for your invaluable friendship over the years and future!

Thank you Marian for your precious friendship over the years and delightful company. Thanks to my friends in India, Starkville, and other countries—for your friendship through the years—my life wouldn't be complete without you. A distinct note of thanks to Anne (Granny), Ted, and Mr. Bush for your company, love and service to MSU shuttle and student community.

A noteworthy appreciation to all the speakers and talks that I got to attend for professional development, and all the amazing women role models out there. You have all touched my life and helped me grow both personally and professionally. Kudos to your inspiring work!

Most importantly and above all, with all my heart and soul, Thank you God for your guidance, protection and love. I wouldn't be here without you!

Let's all achieve greatness together and strive towards a better tomorrow! Cheers to our bright future and long live humanity!

– Vineetha Menon

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

The advent of hyperspectral imagery has propelled many applications in the area of remote sensing for earth observation because of the vast spectral information contained in its dense, contiguous reflectance bands spread throughout the electromagnetic spectrum [8, 24, 61, 62, 85, 86]. However, this enormous spectral information poses challenges to analysis tasks—due to the well-known *curse of dimensionality* [55]—in addition to incurring high storage, computation, and transmission costs. Dimensionality reduction is the most commonly used tool to mitigate the aforementioned issues. Generally, dimensionality-reduction techniques are used to project high-dimensional data to a low-dimensional subspace based on some objective function and with the goal that all the important information in the data is preserved. The assumption is that the corresponding low-dimensional data has decorrelated bands and lower redundancy of information, thereby aiding classification of land covers [6, 11, 13, 34, 37, 45, 46]. Widely used dimensionality-reduction techniques include singular value decomposition (SVD) [59], principal component analysis (PCA) [40, 44, 58], linear discriminant analysis (LDA) [7, 38], as well as local Fisher discriminant analysis (LFDA) [93] and its kernel variants [10, 25, 39, 51].

Typically, transform-based dimensionality-reduction techniques provide a data-dependent representation and are computationally expensive due their being driven by learning based on an objective function. Hence, there has been increasing interest towards data-independent dimensionality reduction such as random projections. Not only do random projections provide a data-independent representation, they also are computationally lightweight and simple to implement [1, 9, 17, 22, 23, 27]. Recent efforts have investigated computationally fast and efficient random-projection algorithms. This has consequently led to the introduction of random projections in the form of a Hadamard matrix (HM) as an alternative to more traditional random projections based on a random Gaussian matrix (GM) [1, 2, 47, 75, 82].

This dissertation explores computationally efficient dimensionality reduction for hyperspectral imagery driven by random projections. First, we overview the prior use of random projections for dimensionality reduction in the literature, focusing specifically on HM- and GM-based dimensionality-reduction methods for hyperspectral data [20, 21, 41, 42, 77, 83, 94] and also discussing the theoretical soundness and computational efficiency of such methods. Next, we introduce a new paradigm for dimensionality reduction which couples HM- and GM-based projections with feature selection via fast SVD (FSVD), evaluating the resulting classification performance in the framework of a support-vector-machine (SVM) classifier with radial-basis-function (RBF) kernel [5, 19, 26, 50, 90, 95]. Finally, we propose a new framework that addresses the commonly encountered spectral-unmixing phenomenon in hyperspectral imagery by introducing HM- and GM-based nonnegative least

squares (NNLS) that combines dimensionality reduction with estimation of endmember abundances.

1.2 Contributions

In this dissertation, we present three significant contributions for dimensionality reduction and classification of hyperspectral data. Firstly, we introduce HM as an alternative to GM for random projections for the dimensionality reduction of hyperspectral imagery. While random projections have been used rather extensively in recent literature for the dimensionality reduction of hyperspectral imagery (e.g., [35, 36, 44, 46, 65, 67, 71]), such prior work has focused exclusively on GM-based random projections. Additionally, while random HM projections have been used previously in other domains, the work we present here is, to the best of our knowledge, the first use of HM-based random projections for hyperspectral imagery. HM-based projections are theoretically sound and computationally faster than traditional GM projections because of their unique block-based structure [2, 3] that replaces the expensive matrix-multiplication operation associated with GM projections with a series of additions and subtractions. For hyperspectral imagery, we find that HM-based random projections are not only computationally faster but also give better classification performance than traditional GM-based projections even with far fewer dimensions. This dissertation presents HM-based random projections as a viable candidate for dimensionality reduction. Using HM as an alternative to GM is relevant, particularly when the goal is object classification or identification in a lower-dimensional subspace, rather than reconstruction from the lower-dimensional projections (a common focus of prior lit-

erature). We note that our work on HM-based random projections for the dimensionality reduction of hyperspectral imagery was initially published as [75].

Secondly, this dissertation investigates dimensionality reduction via random projections in conjunction with feature selection using a fast variant of SVD known as FSVD. Feature selection is effectively dimensionality reduction that removes statistically ill-conditioned features and redundancy of information to improve classification performance. As discussed in [18], transform-based dimensionality reduction such as SVD can be computationally cumbersome when dealing with large volumes of data. Thus, there has been an increasing need for transform-based feature-selection/dimensionality-reduction techniques deployed in the random-projection domain in order to substantially reduce computational complexity. FSVD has been proven to reduce computational burdens by providing a good approximation to SVD in the random-projection domain by enabling selection of the desired number of eigenvectors for further dimensionality reduction [18]. The contribution we present here is the incorporation of HM-based projections into the FSVD framework (i.e., HM-FSVD), the original FSVD formulation in [18] using GM-based projections exclusively (i.e., GM-FSVD). Thus, the process we propose here employs a two-stage dimensionality reduction: first, projection with HM-based random projections and, second, feature selection in the HM-projected domain using SVD to select the desired number of eigenvectors to be retained. Experiment results validate that random projections in conjunction with FSVD give better classification performance along with reduction in both time and space complexity in comparison to SVD performed on the original dataset. Addi-

tionally, the proposed HM-FSVD gives classification performance superior to GM-FSVD even at low dimensions. We note that our HM-FSVD work was initially published as [74].

Lastly, we address the spectral-unmixing problem in hyperspectral imagery using NNLS in conjunction with random-projection-based dimensionality reduction. Although hyperspectral imagery provides rich spectral information, it usually comes at the price of low spatial resolution. This leads to the occurrence of "mixed pixels," wherein—due to the large footprint of hyperspectral sensors—a single pixel may, in fact, comprise multiple landcovers and, as a result, can be expressed as a combination of one or more constituent endmembers [32]. To simultaneously address these issues of dense spectral information—which leads to high storage costs—and the "spectral mixing" phenomena, we employ dimensionality reduction driven by random projections followed by NNLS-based spectral unmixing in the low-dimensional space. Dimensionality reduction has many advantages—such as reduced signal disparity, decorrelation of bands, and inherent denoising—that can be especially beneficial for spectral unmixing, yielding more accurate abundance estimation of endmembers. Our contribution in this arena is the coupling of NNLS—typically deployed in the original full-dimensionality space of the hyperspectral pixel—with dimensionality reduction driven by random projections. In this effort, we consider both our proposed HM- and HM-FSVD-based projections. Experimental evaluation demonstrates that, again, the HM-based dimensionality reduction methods give more accurate abundance estimation and superior classification performance with low errors at fewer dimensions as compared to using the original data without dimension reduction. We note that our NNLS-based work was initially published as [76].

The remainder of this dissertation is organized as follows. In Chapter 2, we discuss background relevant to random projections and their general use for dimensionality reduction, introducing the computationally efficient HM-based projections as an alternative to traditional GM-based methods. Next, in Chapter 3, we present our HM-FSVD dimensionality reduction which couples HM-based random projections with FSVD-based feature selection. Then, in Chapter 4, we present our coupling of random projections with NNLS for low-complexity abundance estimation and classification. Finally, Chapter 5 presents concluding remarks and summarizes observations made in this dissertation.

CHAPTER 2

BACKGROUND

2.1 Dimensionality Reduction Using Random Projections

As already mentioned, the rich spectral information contained in hyperspectral imagery has given rise not only to widespread applications in remote sensing, but also several impediments to its use. These latter issues include the *curse of dimensionality* [55]—which occurs when the number of training samples is less than the number of spectral dimensions—as well as other issues such as high storage and computation costs. To overcome these problems, dimensionality reduction plays a crucial role for facilitating data analysis in low-dimensional spaces through various forms of feature reduction and feature selection/extraction [87]. Furthermore, dimensionality reduction aids classification by decorrelating spectral bands, reducing redundancy of information, and providing inherent denoising. Some widely used dimensionality-reduction techniques include principal component analysis (PCA) [40, 58, 72, 80] as well as linear discriminant analysis (LDA) [7, 38] and its variants [25, 52, 93].

Generally, transform-based dimensionality reduction—such as PCA and LDA—is designed so as to optimize an objective function aimed at learning the underlying structure of the data. Consequently, transform-based methods are highly data dependent and computationally intensive, especially when dealing with the large volumes of data typ-

ically associated with hyperspectral imagery. On the other hand, dimensionality reduction based on random projections has garnered recent interest due to its data-independent representation as well as its ability to preserve important information present in the data [9, 17, 43, 48, 60, 92]. Dimensionality reduction via random projection is based on the assumption that a valid and useful low-dimensional representation can be had even when the low-dimensional space is chosen at random. Since there is no data-specific learning process involved, random projections are data independent and computationally lightweight. Ultimately, this simplicity of implementation and computational efficiency of random-projection-based methods makes them a favorable candidate for dimensionality reduction of hyperspectral data [1, 28, 36, 45, 47, 53, 57, 64, 69]. This chapter of this dissertation focuses on applications of fast random projections for dimensionality reduction to address the high computation and storage costs associated with hyperspectral imagery. Specifically, this work explores projections based on a random Hadamard matrix (HM) as an alternative to more traditional random projections which are commonly driven by a random Gaussian matrix (GM). In the rest of this chapter, we first discuss the traditional GM-based approach to random projections in Section 2.2.1 followed by our main focus in Section 2.2.2, HM-based random projections. The material in this chapter was initially published in [75].

2.2 Random Projections

2.2.1 Random Projection Based on a Gaussian Matrix

Let a dataset of M N-dimensional hyperspectral pixels be $\mathbf{X} = \{\mathbf{x}_m\}_{m=1}^{M} \in \mathbb{R}^{N \times M}$; the corresponding class labels are $\mathbf{\Theta} = \{\theta_m\}_{m=1}^{M}$, where $\theta_m \in \{1, 2, \ldots, C\}$, and C

denotes number of classes. Consider projection matrix $\mathbf{P} \in \mathbb{R}^{N \times K}$, where K is the reduced dimension and $K \ll N$. We desire that $\mathbf{P}$ preserves important information present in the data with very high probability and retains data separability in the corresponding low-dimensional space. An increasingly common approach is to randomly choose a $\mathbf{P}$ by populating a matrix with independent random variables (e.g., uniform, Gaussian). In this case, the dimensionality-reduced data matrix $\hat{\mathbf{X}}$ is

$$\hat{\mathbf{X}} = \mathbf{P}^T \mathbf{X}, \tag{2.1}$$

where $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_m\}_{m=1}^{M} \in \mathbb{R}^{K \times M}$. Most commonly, Gaussian random variables are used to populate $\mathbf{P}$, resulting in it being a random GM.

Figure 2.1 illustrates an example of such a GM projection matrix, $\mathbf{P}_{\text{GM}}$, where mid-range gray values are zero, darker values are negative, and lighter values are positive. It has been shown that GM-based random projection achieves data reduction while preserving important information in the corresponding low-dimensional space (e.g., [3]). Traditional GM-based random projections (i.e., $\mathbf{P} = \mathbf{P}_{\text{GM}}$ in (2.1)) require $\mathcal{O}(NMK)$ time. The primary disadvantage of GM-based projections is the heavy computation load imposed by the dense matrix multiplication operations implied by the GM.

2.2.2 Random Projection Based on a Hadamard Matrix

Growing interest in random projections has led to exploration of non-traditional and computationally faster projection matrices (e.g., [2, 3, 75, 81, 82]) based on a random

Hadamard projection. Such projections are driven by a HM defined recursively for any N that is an integer power of 2 as

$$\mathbf{H}_N = \begin{bmatrix} \mathbf{H}_{\frac{N}{2}} & \mathbf{H}_{\frac{N}{2}} \\ \mathbf{H}_{\frac{N}{2}} & -\mathbf{H}_{\frac{N}{2}} \end{bmatrix} \in \mathbb{R}^{N \times N}, \tag{2.2}$$

with $\mathbf{H}_1 = [1]$, which we normalize as $\bar{\mathbf{H}}_N = \sqrt{\frac{1}{N}}\mathbf{H}_N$. Following [31, 81, 82], we define a random diagonal matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$ with diagonal entries being ± 1 with equal probability (i.e., $\frac{1}{2}$). The purpose of this diagonal matrix $\mathbf{D}$ is to randomize $\mathbf{H}_N$. Broadly, a randomizer matrix can be classified as a *global* or *local* randomizer based on the structure of its matrix and the resulting effects on permuted samples. Typically, a uniform random matrix randomizes samples globally, whereas a diagonal random matrix, as in our case, randomizes samples locally [31]. We further define a random sampling matrix $\mathbf{S} \in \mathbb{R}^{N \times K}$ that randomly samples K columns of $\mathbf{D}\bar{\mathbf{H}}_N$, where each column of $\mathbf{S}$ is randomly selected with replacement from the $N \times N$ identity matrix $\mathbf{I}_N$. Thus, the HM-based projection matrix is formulated as

$$\mathbf{P}_{\text{HM}} = \sqrt{\frac{N}{K}}\mathbf{D}\bar{\mathbf{H}}_N\mathbf{S} \in \mathbb{R}^{N \times K}. \tag{2.3}$$

We note that this definition requires that N be an integer power of 2. Consequently, when applying $\mathbf{P}_{\text{HM}}$ as the projection matrix in (2.1), we zero-pad the rows of $\mathbf{X}$ as required such that this condition is satisfied. We further note that $\mathbf{P}_{\text{HM}}$ is orthonormal, and it has been shown to preserve important details in the data in a low-dimensional space [2]. Due to the block-based structure of the HM, each N has $\log N$ "recursion" stages of computation; i.e., for $N = 16$, there are $\log N = \log_2 16 = 4$ stages of "recursion" computation (which are $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_4$, and $\mathbf{H}_8$). However, for dimensionality reduction, we need to compute only

K stages, along with a linear-time merge step in (2.2). Therefore, HM-based projections can be computed by combining (2.2) and (2.3) in (2.1) (using $\mathbf{P} = \mathbf{P}_{\text{HM}}$) in $\mathcal{O}(NM \log K)$ time [81, 82]. This can be significantly more computationally efficient than GM-based random projections that require $\mathcal{O}(NMK)$ time. Furthermore, since a HM comprises ± 1 entries, the matrix multiplication in (2.1) reduces to a series of additions and subtractions. Figure 2.2 illustrates an example HM, $\mathbf{H}_N$ for $N = 2^i$, where $i \in 0, 1, \ldots, 4$, where $white$ represents $+1$ values, and $black$ represents -1 values. Figure 2.3 shows the HM-based-projection matrices $\mathbf{P}_{\text{HM}}$ for same range of N, where in this case $white$ denotes $+\sqrt{\frac{N}{K}}$, and $black$ denotes $-\sqrt{\frac{N}{K}}$. The effects of the random diagonal matrix $\mathbf{D}$ in randomizing $\bar{\mathbf{H}}_N$ can been seen in Figure 2.3.

We can draw similarities both computationally and structurally between the Hadamard Transform (HT) and Discrete Fourier transform (DFT) matrix, denoted by the matrix $\mathbf{F}_N$ given as

$$\mathbf{F}_N = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & & \omega^{(N-1)} \\ \vdots & & & \ddots & \vdots \\ 1 & \omega^{(N-1)} & \omega^{2(N-1)} & \ldots & \omega^{(N-1)^2} \end{bmatrix} \in \mathbb{R}^{N \times N}, \tag{2.4}$$

where $\omega = e^{-\frac{i2\pi}{N}}$ implies complex-valued entries in $\mathbf{F}_N$. Both $\mathbf{F}_N$ and $\mathbf{H}_N$ are similar in terms of their computational complexity due to their block-based structure (as in (2.2) and (2.4)). For example, consider $N = 2$. In this case, the HM is

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_1 \\ \mathbf{H}_1 & -\mathbf{H}_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{2.5}$$

which is identical to the DFT matrix

$$\mathbf{F}_2 = \begin{bmatrix} 1 & 1 \\ 1 & \omega^{(2-1)} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{2.6}$$

For higher orders of N, $\mathbf{H}_N$ still comprises ± 1 terms, whereas the DFT matrix $\mathbf{F}_N$ is composed of complex-valued entries, such that $\mathbf{H}_N$ and $\mathbf{F}_N$ diverge in structure and meaning, the DFT being well-understood in terms of an expansion of sinusoids. On the other hand, HT is a generalized class of DFT, and is in fact equivalent to a multidimensional DFT of size of power of 2 [**?**].

Figure 2.4 illustrates the block-based structure of both the real and imaginary components of $\mathbf{F}_N$ for varying values of N (scaled such that white is $+1$, black is -1), which is seen to differ from the corresponding HM example (Figure 2.3).

In the following chapters, we deploy the Hadamard-based random projections we describe in this section for achieving data-independent dimensionality reduction with reduced computational complexity as compared to the traditional Gaussian-based projections that have been used extensively in prior literature. In the next chapter, we specifically employ HM-based random projections in conjunction with a fast singular-value feature selection to implement a two-stage dimensionality-reduction process. Then, in Chapter 4, we couple HM with nonlinear least squares to achieve spectral unmixing in a reduced-dimensionality space.

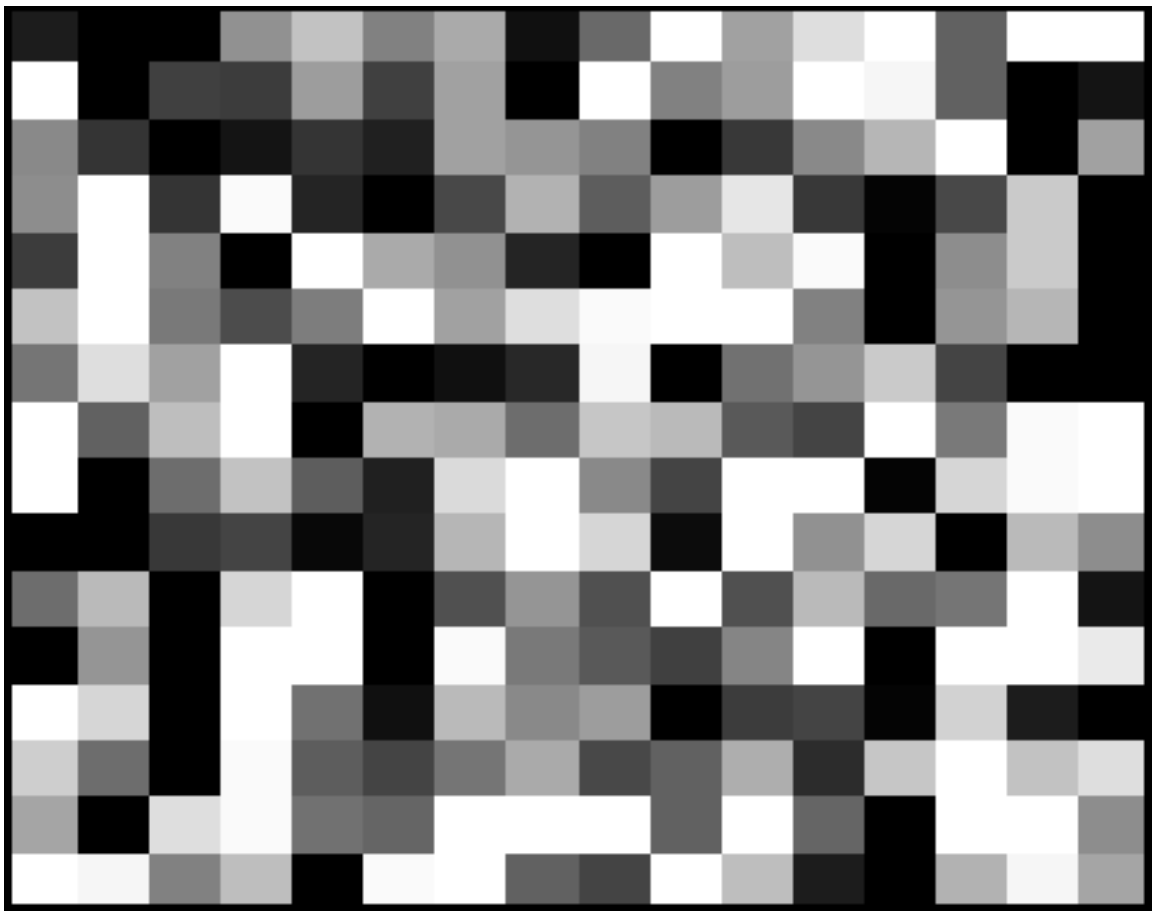
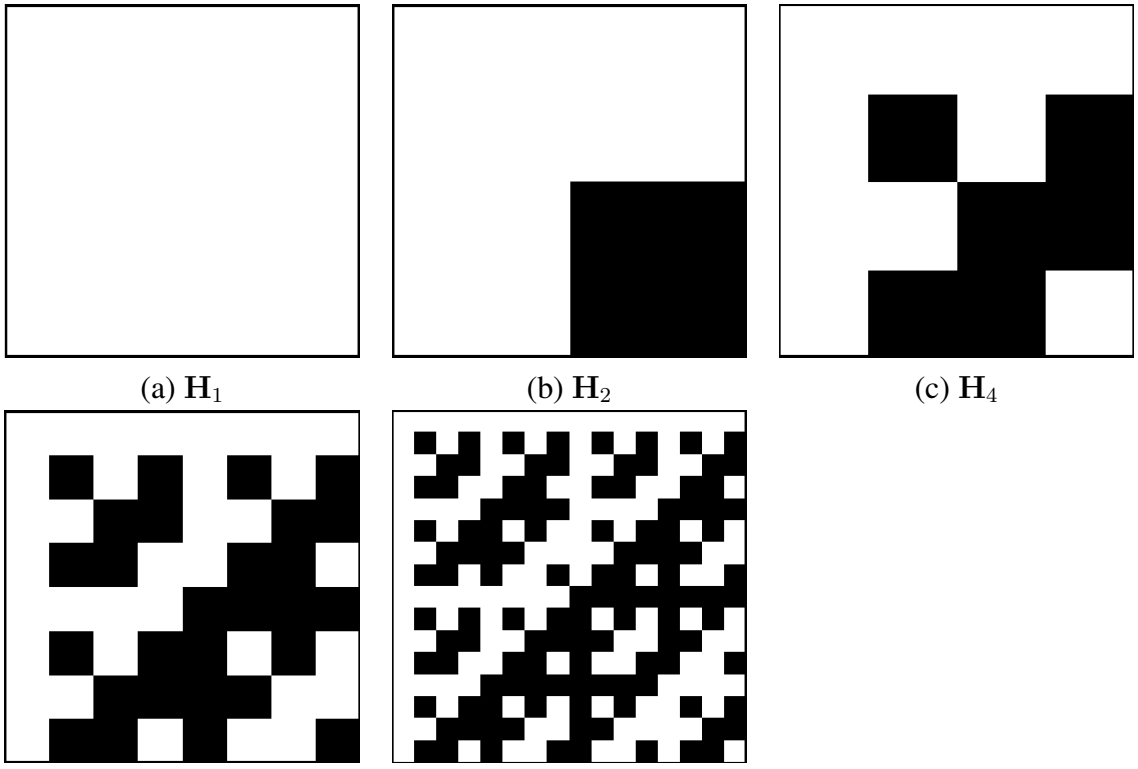


Figure 2.1

Gaussian random-projection matrix $\mathbf{P}_{\text{GM}}$ for $N = 16$

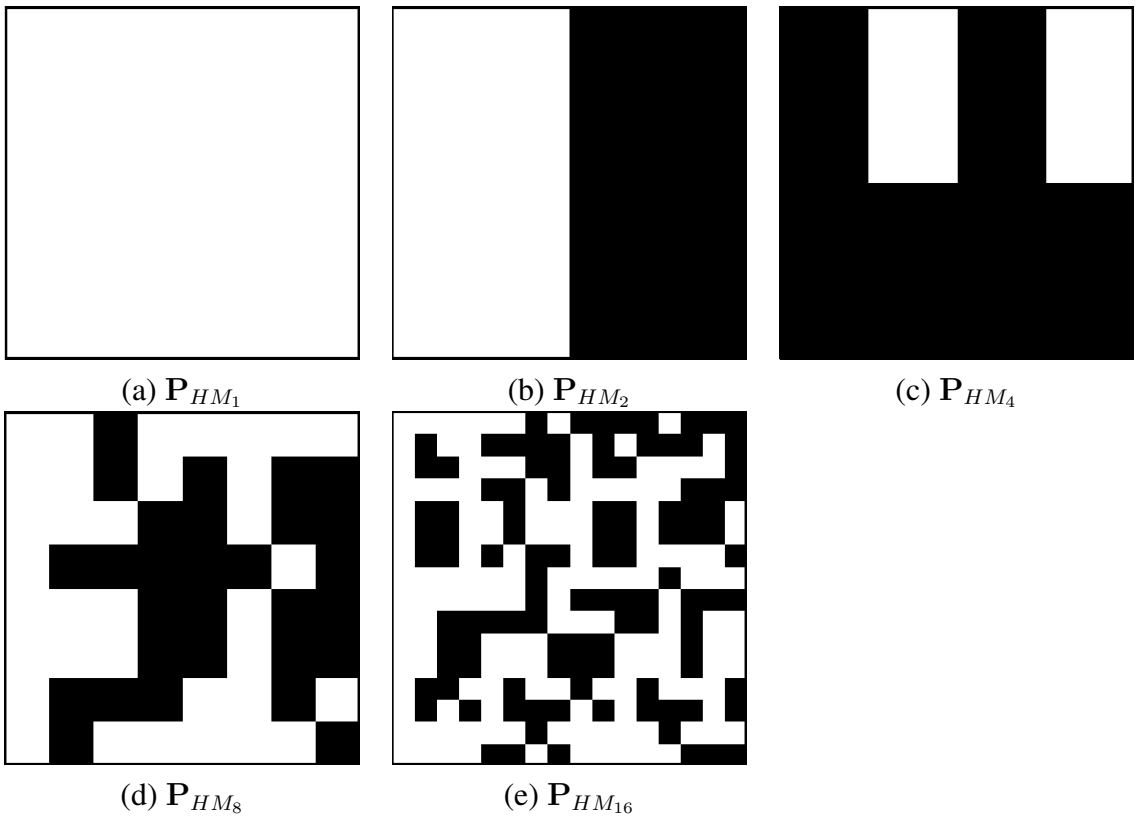


(a) $\mathbf{H}_1$ (b) $\mathbf{H}_2$ (c) $\mathbf{H}_4$

(d) $\mathbf{H}_8$ (e) $\mathbf{H}_{16}$

Figure 2.2

Hadamard matrix $\mathbf{H}_N$

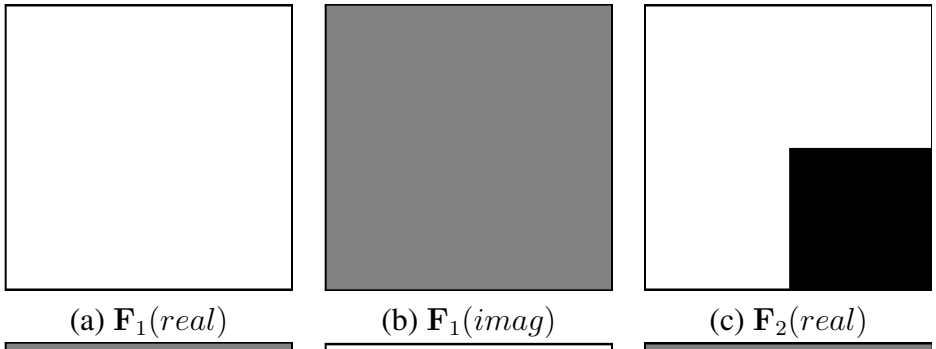


(a) $\mathbf{P}_{HM_1}$ (b) $\mathbf{P}_{HM_2}$ (c) $\mathbf{P}_{HM_4}$

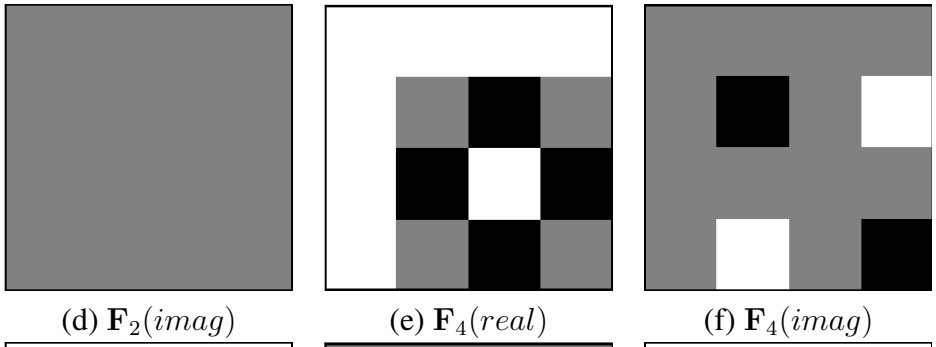
(d) $\mathbf{P}_{HM_8}$ (e) $\mathbf{P}_{HM_{16}}$

Figure 2.3

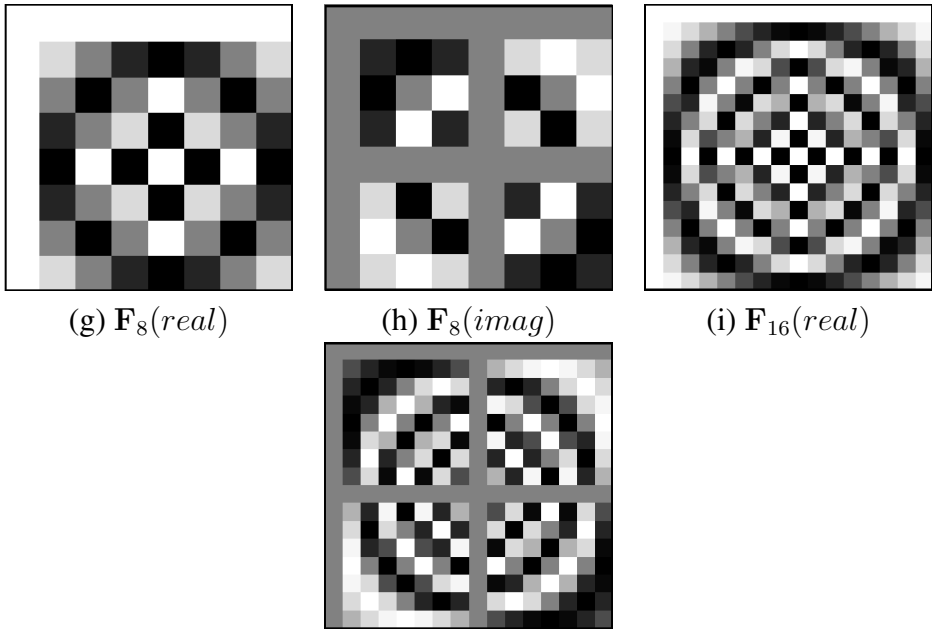
Hadamard random projection matrix $\mathbf{P}_{\text{HM}}$



(a) $\mathbf{F}_1(real)$ (b) $\mathbf{F}_1(imag)$ (c) $\mathbf{F}_2(real)$



(d) $\mathbf{F}_2(imag)$ (e) $\mathbf{F}_4(real)$ (f) $\mathbf{F}_4(imag)$



(g) $\mathbf{F}_8(real)$ (h) $\mathbf{F}_8(imag)$ (i) $\mathbf{F}_{16}(real)$

(j) $\mathbf{F}_{16}(imag)$

Figure 2.4

Real and imaginary components of DFT matrix $\mathbf{F}_N$

CHAPTER 3

FAST SVD WITH RANDOM HADAMARD PROJECTION FOR HYPERSPECTRAL DIMENSIONALITY REDUCTION AND CLASSIFICATION

3.1 Introduction

As discussed previously in this dissertation, many popular approaches to the dimensionality reduction of hyperspectral imagery take the form of data-dependent transforms. Most instances of such transform-based dimensionality reduction involve mapping high-dimensional data to a lower-dimensional subspace based on some objective function yielding an efficient yet data-dependent representation. Singular value decomposition (SVD) [59]—and the closely related principal component analysis (PCA) [40, 44, 58]—are common examples of data-dependent transform-based dimensionality reduction that are highly effective in many scenarios, yet their data dependent nature often entails substantial computational burden.

In this chapter, we explore several strategies for dimensionality reduction and feature selection driven by random projections to reduce the time and space complexity involved when dealing with hyperspectral imagery. We consider both the more traditional GM-based random projection as well as the faster HM-based counterpart, both of which were introduced in Chapter 2 of this dissertation. In this chapter, we focus on coupling both GM- and HM-based dimensionality reduction with SVD-driven feature selection to effectively

produce a two-stage dimensionality reduction. At first glance, data-independent random projections may seem diametrically opposed in spirit to data-dependent techniques—such as SVD—that optimize the dimensionality reduction to the dataset at hand; however, recent work (e.g., [18, 85]) has suggested that the two strategies can effectively complement one another. For example, [18] proposes a fast SVD driven by Gaussian-based random projection that enables finding an approximation to the SVD dimensionality-reduction operator with dramatically reduced computational complexity. Such a fast SVD (FSVD) yields a surprisingly effective proxy for the true SVD at a fraction of the computational cost.

In experimental results presented in this chapter, we explore the effect of dimensionality reduction on the classification performance of hyperspectral imagery using both a supervised support-vector-machine (SVM) classifier as well as an unsupervised k-means clustering. We employ both random projections alone as well as in conjunction with FSVD. We find that, while random projections alone offer extremely fast dimensionality reduction, the coupling with SVD in the form of FSVD offers an attractive tradeoff between classification performance and reduced computation costs with the added advantage of an efficient data-dependent representation at low-dimensions as a close approximation to the original data space. We experimentally validate the data-preserving property of random-projection-based dimensionality reduction and empirically prove the computational efficiency of FSVD over SVD. We note that random projection, as a form of dimensionality reduction, is not intended to enhance class separability. To optimize classification accuracy, a projection designed specifically for that purpose is warranted; such a projection would imply labeled samples and a data-dependent projection. We emphasize that, in contrast,

our goal here is to achieve a tradeoff between classification performance and reduced computational complexity. Consequently, we advocate the proposed Hadamard-based random projection for FSVD due to its lower computational cost as well as classification accuracy that is superior to that of the other schemes considered.

The remainder of this chapter is organized as follows. In Section 3.2, we describe FSVD as a fast and efficient approximation of SVD, but with lower computation complexity than SVD performed in original data space. As our primary contribution, we couple FSVD with the HM-based random projections that were described in Chapter 2. Then, in Section 3.3, we experimentally validate the efficacy of our proposed approaches. We note that the work presented in this chapter was initially published as [74, 75].

3.2 FSVD

Singular value decomposition (SVD), is one of the most frequently used tools for dimensionality reduction and feature selection. In SVD, we express any given matrix in terms of its eigenvectors and eigenvalues and perform dimensionality reduction by choosing the desired number of eigenvectors from those with the largest-magnitude eigenvalues in an effort to preserve important information in the matrix. For any given matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, its SVD can be computed as

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{3.1}$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ contains the left singular vectors (orthonormal eigenvectors), $\mathbf{\Sigma} \in \mathbb{R}^{N \times M}$ contains the eigenvalues, and $\mathbf{V}^T \in \mathbb{R}^{M \times M}$ contains the right singular vectors (orthonormal eigenvectors). For dimensionality reduction, one retains only the K eigen-

vectors in $\mathbf{U}$ corresponding to the K largest eigenvalues, yielding a matrix $\mathbf{U}_K \in \mathbb{R}^{N\times K}$ and

$$\hat{\mathbf{X}} = \mathbf{U}_K^T \mathbf{X} \in \mathbb{R}^{K\times M}. \tag{3.2}$$

SVD has long been used in many applications due to its conceptual simplicity and its theoretic optimality in the sense of minimizing $\left\|\mathbf{X} - \mathbf{U}_K\hat{\mathbf{X}}\right\|_F$ (the Eckart-Young theorem, e.g., [70]). However, it is computationally expensive, requiring $\mathcal{O}(MN^2)$ time, assuming $M \geq N$.

Recently, it has been proposed [18, 89] to expedite the SVD calculation by applying it subsequent to a random projection. In such a fast formulation of SVD (which we denote as FSVD), the computationally burdensome SVD process effectively takes place in a lower-dimensional space at the expense of being only an approximation to the exact SVD as calculated in the full-dimensional space. We extend the FSVD algorithm from [18] by using an HM-based random projection, as illustrated in Figure 3.1. This FSVD generates a projection matrix by first applying a random projection into R-dimensional space before reducing dimensionality further to K via an SVD-based projection. Assuming $K < R \ll N$, the SVD in Step 4 of Figure 3.1 has computation $\mathcal{O}(NR^2)$—much less than SVD applied directly to the original dataset. Consequently, the $\mathbf{XQ}^T$ computation in Step 4 dominates, meaning that the overall FSVD algorithm (including random projection in Step 2) runs in $\mathcal{O}(NMR)$ time. Below, we denote the proposed FSVD based on a Hadamard random projection as "HM-FSVD" while the original Gaussian-based strategy from [18] is labeled "GM-FSVD."

3.3 Experimental Results

3.3.1 Supervised Classification

We experimentally validate the effectiveness of the various random-projection dimensionality-reduction approaches by evaluating supervised-classification performance using a support-vector-machine (SVM) classifier with a radial-basis-function (RBF) kernel. We use the Indian Pines[1] and University of Pavia [49] datasets in our experiments. The Indian Pines dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Indian Pines test site in northwestern Indiana. The dataset has a spatial dimension of 145×145 with a spatial resolution of 22 m, 224 spectral bands, and 16 land-cover classes. After removal of 22 water-absorption bands, the dataset was reduced to 202 spectral bands. The University of Pavia dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) over an urban area of Pavia in north Italy. This dataset has 103 spectral bands each having a spatial dimension of 610×340 with a spatial resolution of 1.3 m, with 9 classes of land cover. For each dataset we randomly select 10% training and 90% testing data. Tables 3.1 and 3.2 give the ground-truth classes as well as the number of training and testing samples used. All SVM experiments were implemented with libSVM[2].

We now compare five methods of dimensionality reduction, namely: Hadamard- and Gaussian-driven random projection alone (denoted as "HM" and "GM," respectively); HM-FSVD and GM-FSVD as described in Sec. 3.2; and SVD applied directly to the full-dimensionality dataset (denoted as "SVD"). We also compare to the SVM-RBF classifier

[1] `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`
[2] `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

applied to the original dataset with no dimensionality reduction (denoted as "Original"). RBF kernel parameters were chosen after 10-fold cross validation, and all the experiments were run multiple times with average classification accuracy reported. For dimensionality reduction using HM, GM, and SVD, the dataset is reduced directly to its final dimensionality via (2.1) or (3.2); we use a final dimensionality $K \in \{14, 28, 43, 57, 71\}$ for Indian Pines and $K \in \{7, 15, 22, 29, 37\}$ for Pavia.

On the other hand, in case of HM-FSVD and GM-FSVD, the random projection first reduces dimensionality to R and then the SVD-based process reduces dimensionality further to K. We choose $R \in \{20, 40, 61, 81, 101\}$ for Indian Pines and $R \in \{10, 21, 31, 41, 52\}$ for Pavia, which corresponds to a reduction of approximately 10–50% of the original dataset dimensionality. On the other hand, final dimensionality K is the same as used for the other methods; we note that we have chosen the intermediate dimensionality R such that $K = 0.7R$. We have found that, for both HM-FSVD and GM-FSVD, choosing K and R is highly application specific—based on the degree of dimensionality reduction and number of eigenvectors chosen, there can be varying effects on final classification accuracy. If the number of eigenvectors K is too small, then this leads to loss of information in the data, and if K is too large, then it defeats the purpose of dimensionality reduction and sometimes may even cause a drop in classification accuracy due to correlation of spectral bands. Here, for simplicity of presentation, we present results exclusively for $K = 0.7R$.

Figure 3.2 depicts the first three eigenvectors that result from SVD on the original dataset as well as those from the FSVD-based methods. Additionally, Figure 3.3 gives the spectral angle (in degrees) between the eigenvectors produced by the FSVD-based methods

and those from SVD as determined directly from the original dataset with full dimensionality. We argue that the proposed FSVD approaches, despite extracting eigenvectors in the random-projection domain, do indeed produce eigenvectors similar to the more computationally expensive SVD on the original dataset. We observe that HM-FSVD provides closer eigenvectors than does GM-FSVD.

Figs. 3.4 and 3.5 illustrate how classification accuracy varies with the reduced dimensionality, while Tables 3.3 and 3.4 tabulate classification performance for a fixed final dimensionality; Figs. 3.6 and 3.7 show the corresponding classification maps. Finally, computational cost is presented in Tables 3.5 and 3.6 using MATLAB running on a quad-core 3.2-GHz machine with 5.8 GB of RAM.

We see from Tables 3.5 and 3.6 that, as expected, random projections applied alone (i.e., the HM and GM projections) provide the fastest dimensionality reduction, with the Hadamard projection being about an order of magnitude faster than the Gaussian projection, while both are some 3–6 orders of magnitude faster than SVD. However, classification performance for these two random projections is significantly inferior to that of SVD as witnessed in Figs. 3.4 and 3.5 as well as Tables 3.3 and 3.4. Intuitively, this is also as expected, as we tend to view the SVD as the "ideal" dimensionality reduction.

However, the optimality of SVD is merely in the sense of the Eckart-Young theorem; i.e., SVD provides the closest $\|\cdot\|_F$-norm approximation to the dataset for a given reduced dimensionality. Importantly, optimality from this perspective does not necessarily imply optimal classification performance (see, e.g., [84]). To wit, we see in Figs. 3.4 and 3.5 as well as Tables 3.3 and 3.4 that the FSVD-based dimensionality reductions can outperform

SVD even though they provide only an approximation to the true SVD eigenvectors. This is despite their being some 2–5 orders of magnitude faster than SVD.

3.3.2 Unsupervised Clustering

In this section, we study the performance of our proposed methods within an unsupervised-learning paradigm. Specifically, we evaluate the efficacy of our proposed methods using traditional k-means clustering on an AVIRIS dataset, Salinas-A, which is a subsection of the original Salinas dataset acquired over Salinas Valley, California[3]. Salinas-A has a spatial dimension of 83×86 and 224 spectral bands (after removal of 20 water-absorption bands, there are 204 spectral bands) with a high spatial resolution of 3.7 m covering 6 land-cover classes. Table 3.7 gives the number of samples present in each class.

In prior literature, k-means is a common tool for unsupervised classification because of its simplicity and ease of implementation. However, the choice of the number of clusters k can prove to be critical to performance. Many cluster-validation algorithms (such as AIC [4] and BIC [90]) have been developed to estimate the maximum number of clusters present in a given dataset. However, in our case, for simplicity, we assume that the number of clusters is the same as the number of known classes present in the Salinas-A dataset—that is, we set $k = 6$. In order to maintain consistency with the supervised methods evaluated in the previous section, the number of reduced dimensions is set as $K = 0.7R$ with $K \in \{14, 28, 42, 57, 71\}$.

[3] `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`

The same methods considered in Section 3.3.1 are compared, namely, dimensionality reduction performed on the original data using random projections (both HM and GM), random projection with FSVD-based learning (HM-FSVD and GM-FSVD), and dimensionality reduction using SVD on the original data (denoted "SVD") All the above methods are followed by unsupervised learning using k-means clustering; additionally, k-means is applied to the original dataset without any dimensionality reduction (denoted as "Original"). Figure 3.8 shows the clustering accuracy over varying dimensions. As expected, the unsupervised learning methods have uniformly lower classification accuracy as compared to their supervised counterparts. That said, we see that all the methods driven by random projections outperformed the one on the original data.

3.4 Observations

From the results presented in this chapter, we see that HM-FSVD generally outperforms the other dimensionality-reduction strategies for supervised classification, while, for unsupervised clustering, all the random-projection dimensionality-reduction methods perform equivalently. While not as fast as Hadamard-based random projection applied alone, HM-FSVD is generally significantly faster than its Gaussian-based counterpart GM-FSVD while being substantially faster than SVD. Consequently, we conclude that HM-FSVD offers a computationally attractive random-projection-based alternative to SVD for dimensionality reduction in supervised hyperspectral-classification applications. HM-FSVD constitutes a reasonable strategy for dimensionality reduction for unsupervised clustering as well.

Up to this point, this dissertation has considered random projections of hyperspectral imagery with a particular eye towards their effect and interaction with classification processes applied to reduced-dimensional hyperspectral imagery. In next chapter, we turn our attention to another important task in hyperspectral image analysis—the well-known spectral-unmixing problem.

Table 3.1

Ground-truth classes and their samples/class for Indian Pines

Classes			
No.	*Name*	*Train*	*Test*
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-min	83	747
4	Corn	24	213
5	Grass/Pasture	48	435
6	Grass/Trees	73	657
7	Grass/Pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-notill	97	875
11	Soybean-min	246	2209
12	Soybean-clean	59	534
13	Wheats	21	184
14	Woods	127	1140
15	Building-Grass-Trees-Drives	39	347
16	Stone-steel Towers	9	84
Total		1027	9222

Table 3.2

Ground-truth classes and their samples/class for University of Pavia

Classes			
No.	*Name*	*Train*	*Test*
1	Asphalt	663	5968
2	Meadows	1865	16784
3	Gravel	210	1889
4	Trees	306	2758
5	Metal sheets	134	1211
6	Bare soil	503	4526
7	Bitumen	133	1197
8	Bricks	368	3314
9	Shadows	95	852
Total		4277	38499

Table 3.3

Class-specific, overall accuracy (OA), and κ statistics for Indian Pines with $K = 28$

$Class$	**HM**	**HM-FSVD**	**GM**	**GM-FSVD**	**SVD**	**Original**
1	63.41	**70.73**	41.46	58.54	60.98	60.98
2	79.53	**84.28**	74.24	79.92	79.61	79.92
3	68.14	**73.90**	62.65	72.69	71.22	73.76
4	58.68	61.97	59.15	64.32	**64.78**	49.76
5	93.33	**95.17**	93.10	94.25	94.25	93.56
6	**97.56**	96.35	95.74	97.11	97.11	97.41
7	80	84	84	84	84	**88**
8	98.84	99.30	99.53	**99.77**	**99.77**	99.53
9	50	**66.67**	44.44	**66.67**	**66.67**	**66.67**
10	72.11	**84.91**	71.08	76.46	76.57	81.26
11	81.08	86.06	83.30	86.69	**87.10**	83.57
12	73.59	**83.15**	75.47	**83.15**	**83.15**	67.79
13	95.11	96.74	96.20	**97.28**	**97.28**	96.74
14	93.50	96.40	95.60	96.22	**96.92**	95.43
15	56.48	57.92	57.06	**59.37**	**59.37**	53.03
16	83.33	**88.09**	84.52	85.71	85.71	75
OA	82.50	**86.36**	80.17	84.59	84.86	81.96
κ	78.71	**84.14**	78.03	82.62	82.68	80.48

Table 3.4

Class-specific, overall accuracy (OA), and κ statistics for University of Pavia with $K = 15$

$Class$	**HM**	**HM-FSVD**	**GM**	**GM-FSVD**	**SVD**	**Original**
1	90.77	**93.93**	90.33	92.32	92.81	93.62
2	96.50	97.59	97.99	**98.37**	96.37	97.01
3	68.77	**82.21**	62.20	73.90	81.31	79.57
4	89.99	91.33	91.99	92.39	92.20	**93.69**
5	**99.67**	99.59	**99.67**	99.50	95.70	95.70
6	79.96	**88.93**	62.17	82.96	87.58	88.47
7	84.96	**88.80**	70.84	81.62	87.63	87.80
8	85.15	88.41	88.83	**89.62**	83.46	85.48
9	**100**	99.88	**100**	99.88	99.41	99.41
OA	90.98	**93.95**	88.15	92.75	92.41	93.12
κ	87.57	**91.83**	84.96	90.36	89.92	90.86

Table 3.5

Computation time (in seconds) for Indian Pines

K	**HM**	**HM-FSVD**	**GM**	**GM-FSVD**	**SVD**
14	0.002	0.031	0.010	0.062	**2.18**
28	0.001	0.033	0.009	0.067	**2.20**
43	0.001	0.030	0.009	0.067	**2.30**
57	0.001	0.030	0.009	0.064	**2.18**
71	0.001	0.037	0.009	0.062	**2.65**

Table 3.6

Computation time (in seconds) for University of Pavia

K	**HM**	**HM-FSVD**	**GM**	**GM-FSVD**	**SVD**
7	0.002	0.033	0.041	0.072	**3346**
15	0.003	0.042	0.045	0.076	**3819**
22	0.003	0.057	0.043	0.076	**3863**
29	0.003	0.054	0.050	0.082	**3653**
37	0.003	0.084	0.046	0.084	**3751**

Table 3.7

Ground-truth classes and their samples/class for Salinas-A

No.	Class	Samples
1	Brocoli-green-weeds-1	391
2	Corn-senesced-green-weeds	1343
3	Lettuce-romaine-4wk	616
4	Lettuce-romaine-5wk	1525
5	Lettuce-romaine-6wk	674
6	Lettuce-romaine-7wk	799
	Total	5348

1. **Input:** Original data $\mathbf{X} \in \mathbb{R}^{N \times M}$, random projection matrix $\mathbf{P} \in \mathbb{R}^{N \times R}$ (use $\mathbf{P}_{\text{GM}}$ or $\mathbf{P}_{\text{HM}}$ as $\mathbf{P}$)
2. Randomly project $\mathbf{X}$ to R-dimensional space, yielding $\hat{\mathbf{X}} \in \mathbb{R}^{R \times M}$:

$$\hat{\mathbf{X}} = \mathbf{P}^T \mathbf{X}$$

3. Orthonormalize the rows of $\hat{\mathbf{X}}$ to produce $\mathbf{Q}$:

$$\mathbf{Q} = \text{orth}(\hat{\mathbf{X}}) \in \mathbb{R}^{R \times M}$$

4. Perform SVD on $\mathbf{X}\mathbf{Q}^T \in \mathbb{R}^{N \times R}$; i.e.,

$$\mathbf{X}\mathbf{Q}^T = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T$$

5. Let the new projection matrix, $\mathbf{P}_{\text{FSVD}} \in \mathbb{R}^{N \times K}$, contain the largest K singular vectors from $\hat{\mathbf{U}}$ such that $K < R \ll N$.
6. **Output:** New projection matrix $\mathbf{P}_{\text{FSVD}} \in \mathbb{R}^{N \times K}$.

Figure 3.1

The FSVD algorithm for feature selection within the random-projection domain

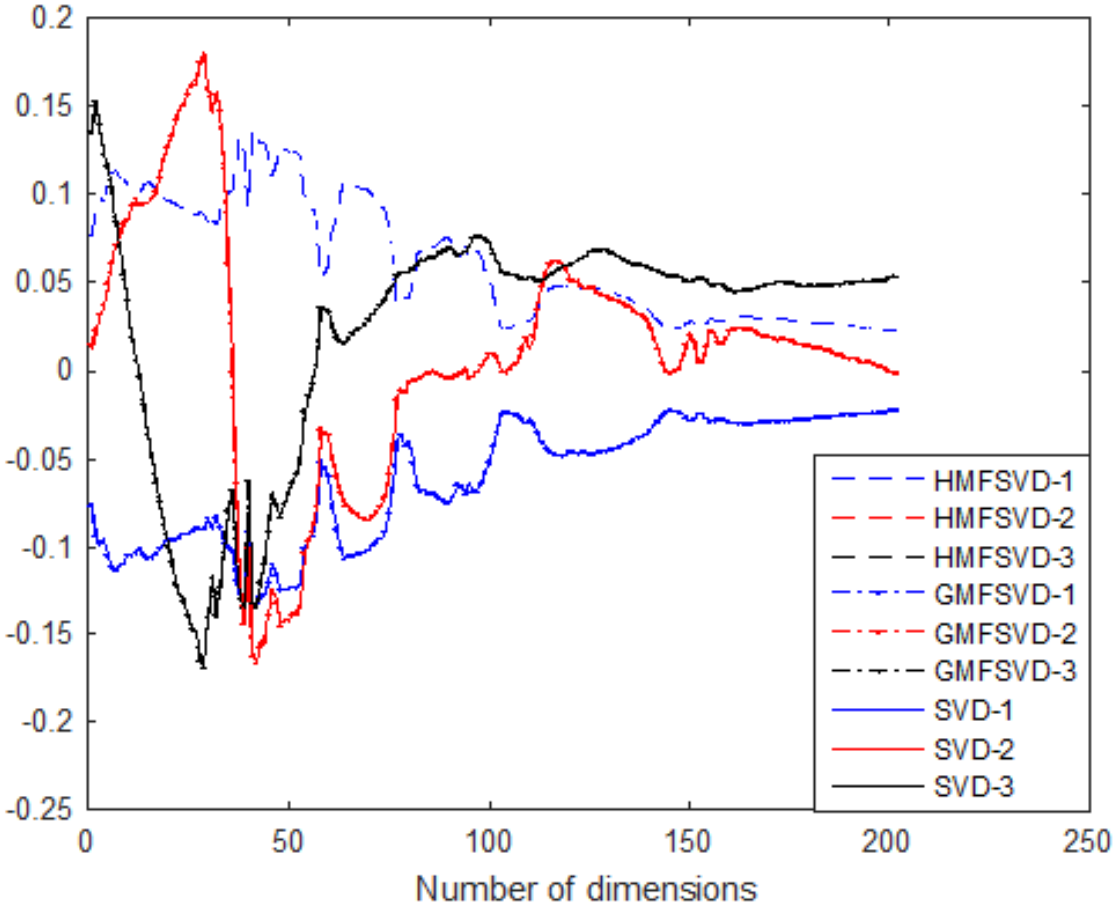


Figure 3.2

The first three eigenvectors as produced by SVD and FSVD-based methods for Indian Pines

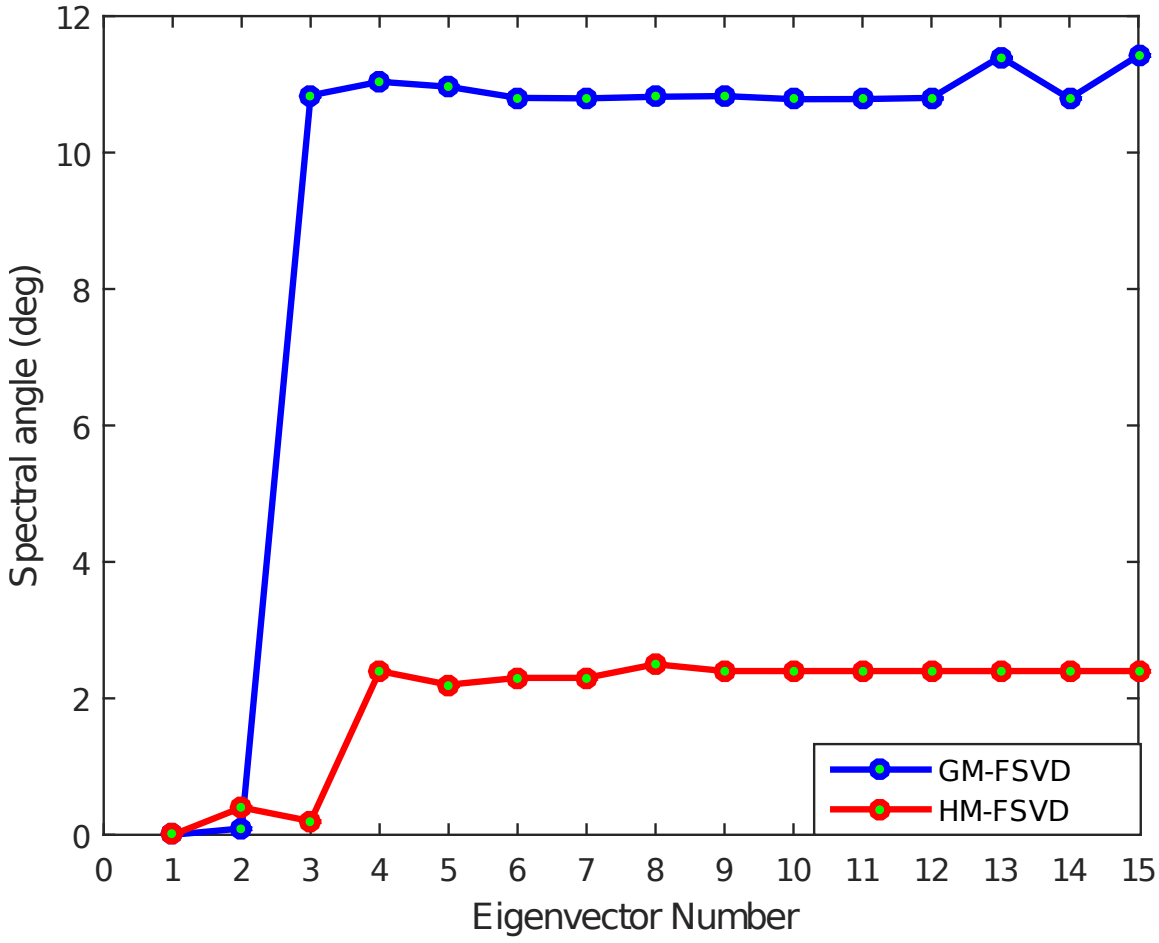


Figure 3.3

Spectral angle between the eigenvectors produced by FSVD-SVD methods for Indian Pines

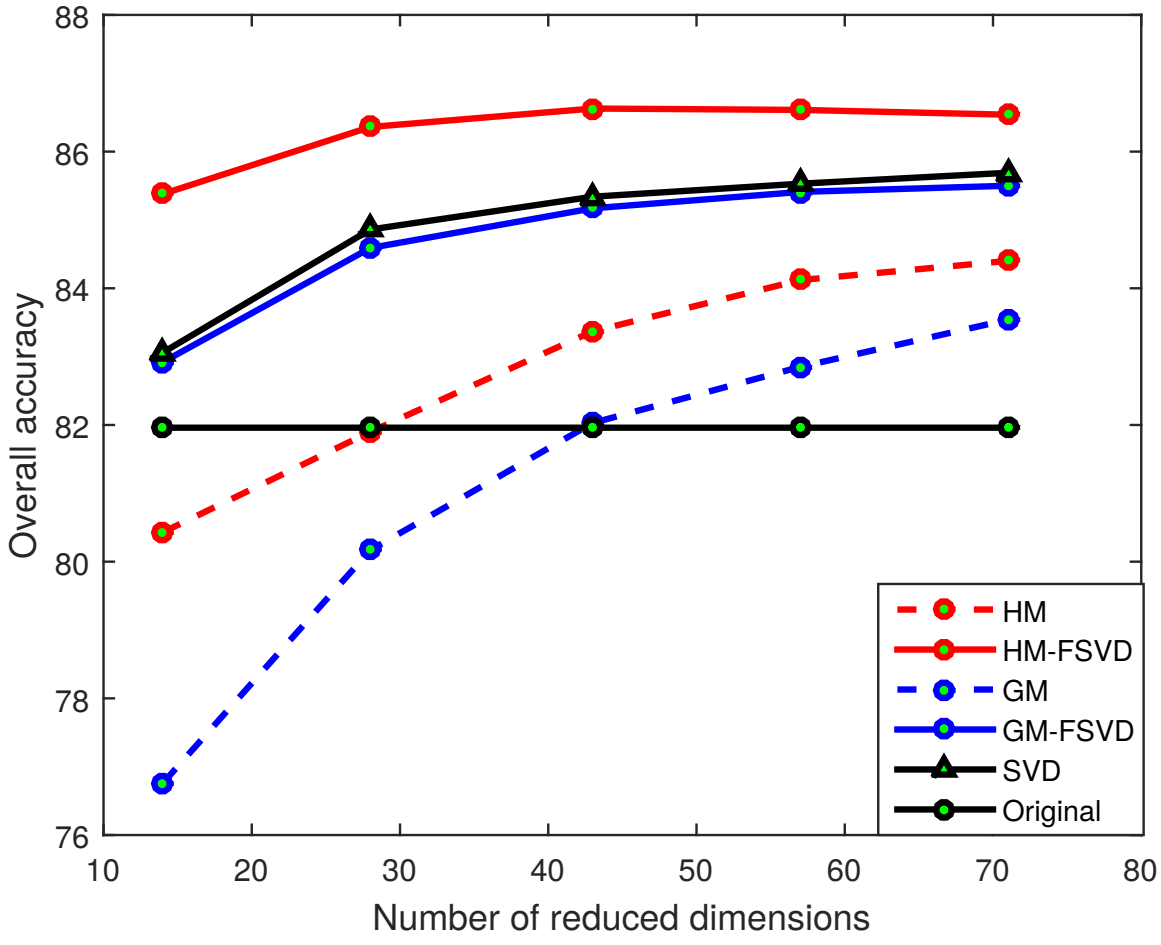


Figure 3.4

Classification accuracy for Indian Pines for varying final dimensionality K

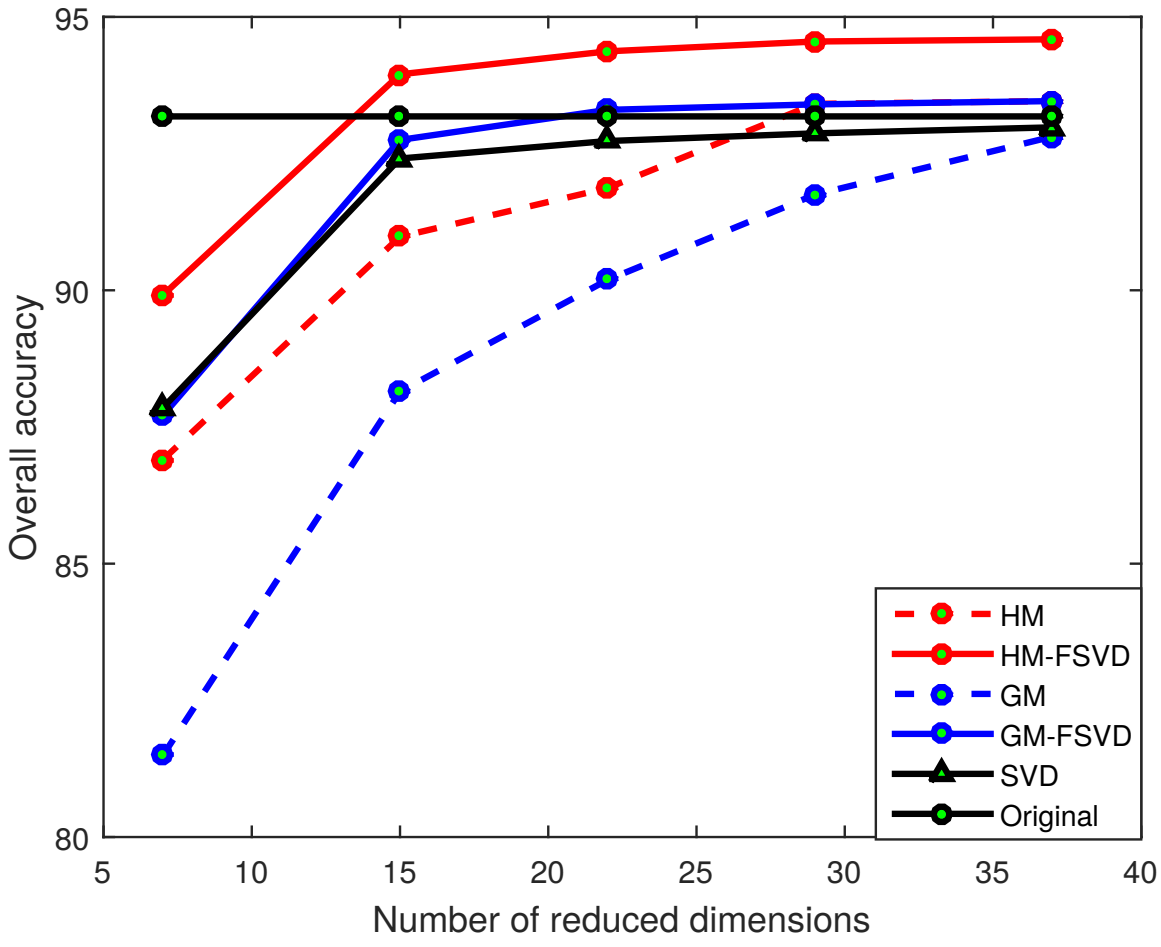
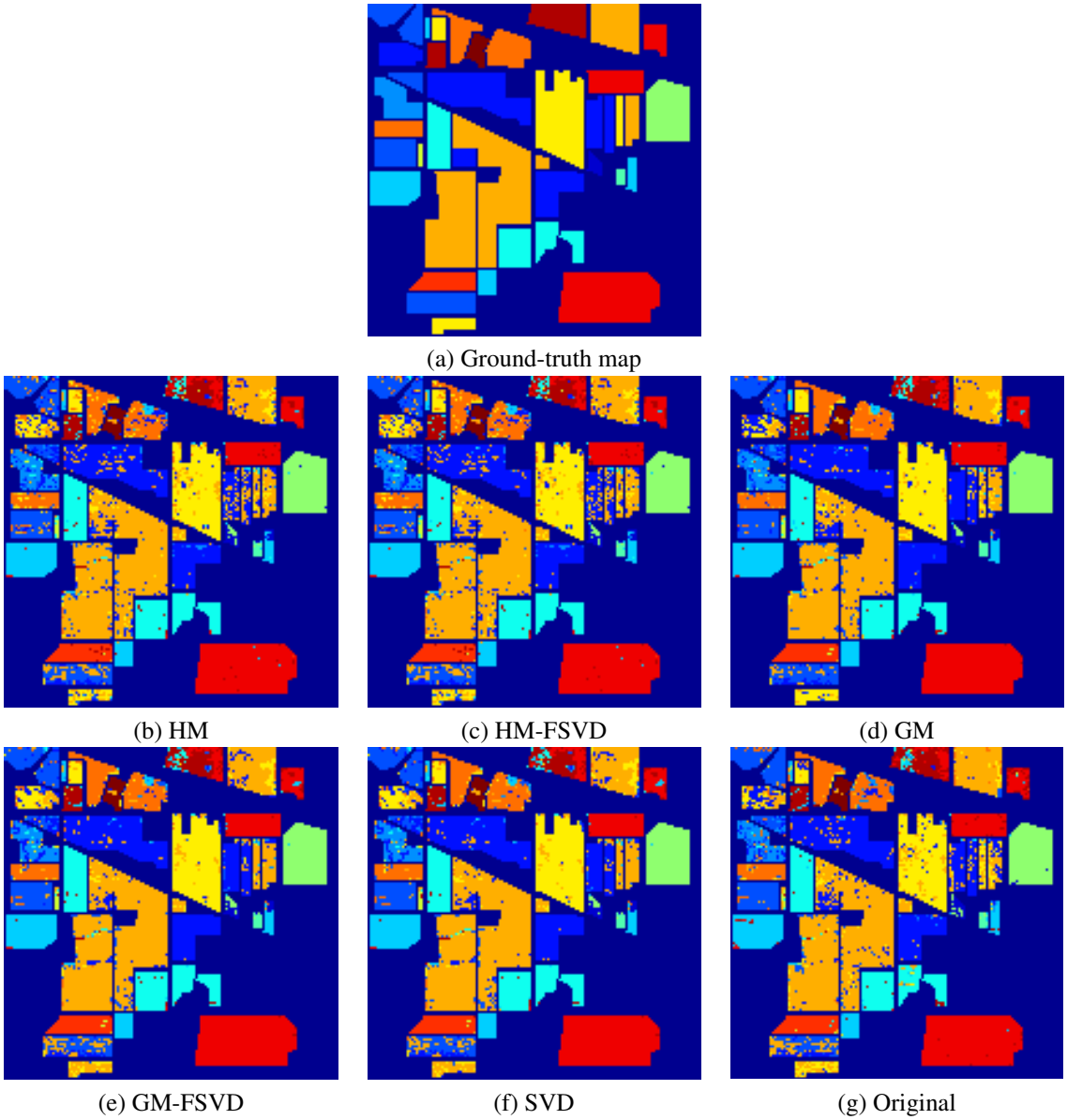


Figure 3.5

Classification accuracy for University of Pavia for varying final dimensionality K



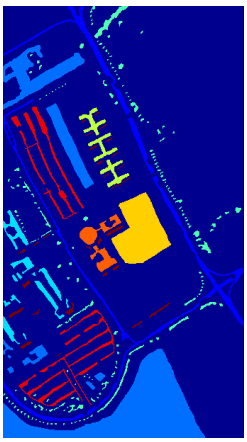
(a) Ground-truth map

(b) HM (c) HM-FSVD (d) GM

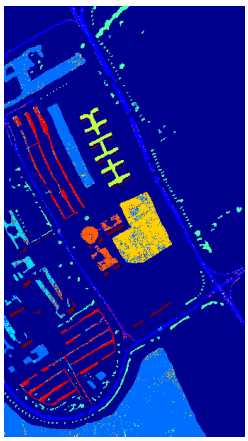
(e) GM-FSVD (f) SVD (g) Original

Figure 3.6

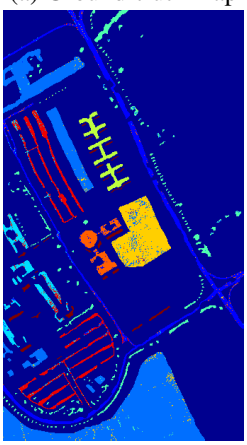
Classification maps for the Indian Pines dataset illustrating different methods using $K = 28$



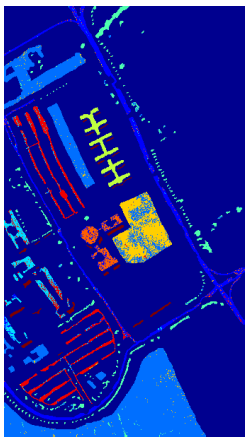
(a) Ground-truth map



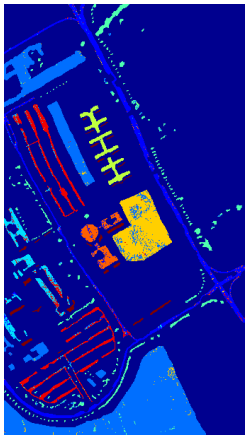
(b) HM



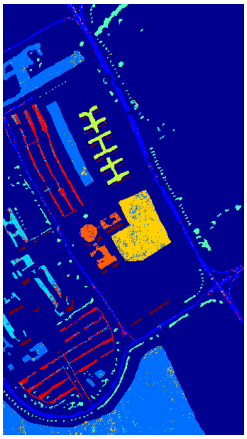
(c) HM-FSVD



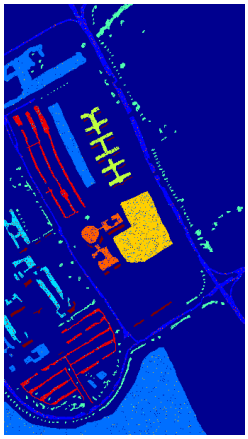
(d) GM



(e) GM-FSVD



(f) SVD



(g) Original

Figure 3.7

Classification maps for the University of Pavia dataset illustrating different methods using $K = 15$

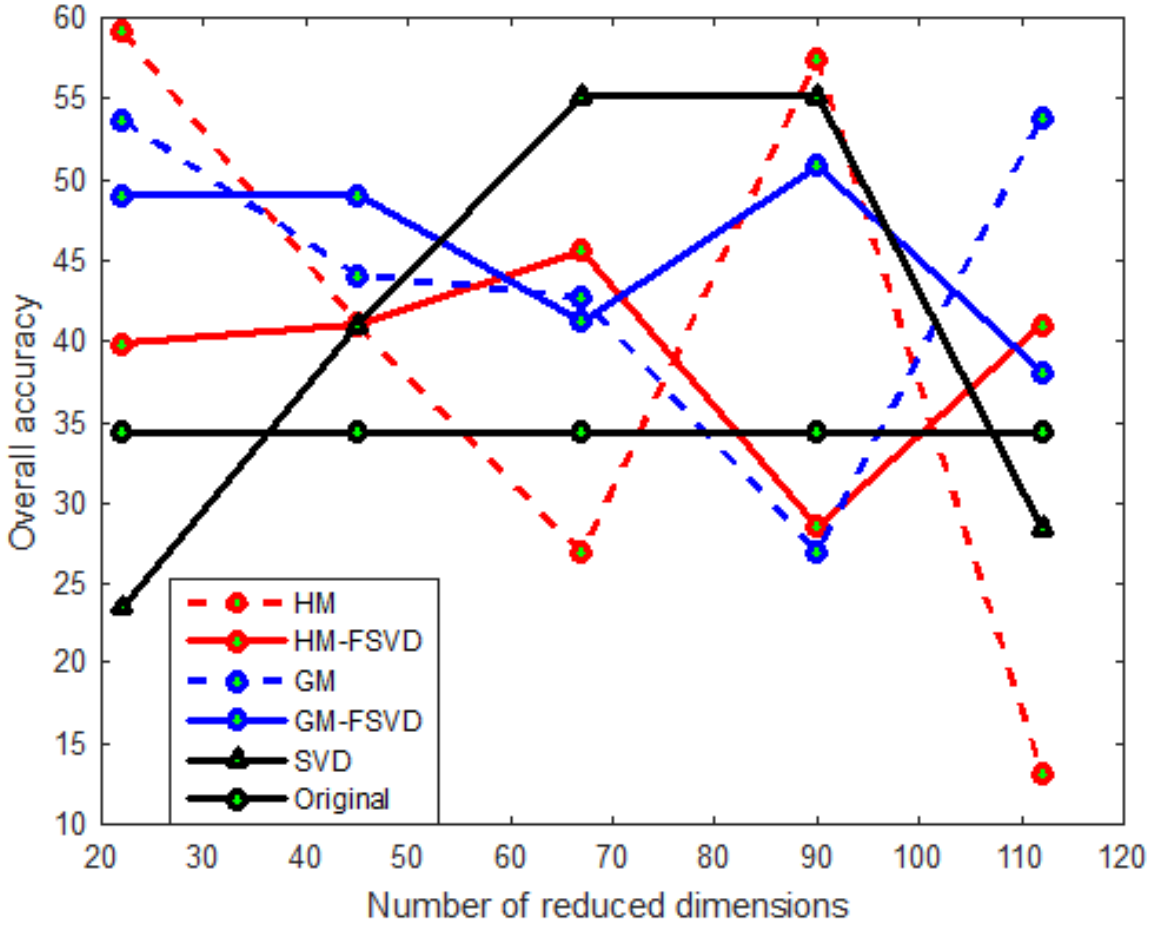


Figure 3.8

Unsupervised clustering performance using k-means on the Salinas-A dataset

CHAPTER 4

RANDOM PROJECTIONS AND NONNEGATIVE LEAST SQUARES FOR SPECTRAL UNMIXING AND CLASSIFICATION

4.1 Introduction

The high spectral resolution of hyperspectral sensors usually comes at the price of low spatial resolution [14]. Due to the large spatial footprint of the sensor, a single pixel typically spans a wide area containing multiple landcover masses, thereby forming a "mixed pixel." Often, this phenomenon is described using a linear mixture model (LMM) wherein the mixed pixel is expressed as a linear combination of its constituent endmembers, the latter being "pure" spectral signatures each describing a single landcover class. In this case, fractional abundances specify the contribution of each endmember to the given mixed pixel. Due to their correspondence to the physical realm, fractional abundances in an LMM must be nonnegative and sum to one in order to describe a mixed pixel realistically. In particular, negative abundance values would have no physical meaning and could not occur in reality.

Prior literature contains numerous methods proposed for the determination, or extraction, of endmember pixels from a hyperspectral image. Some such endmember-extraction techniques include the pixel purity index (PPI) [15], N-FINDR [96], and vertex component analysis (VCA) [79]. Likewise, there exist a multitude of techniques for the esti-

mation of proportional abundances given a set of endmembers; these include constrained optimization algorithms like nonnegative matrix factorization (NMF) [12, 88], nonnegative least squares (NNLS) [12, 16, 56, 68, 78], as well as fully-constrained least squares (FCLS) [54, 73]. The imposition of nonnegativity in these formulations reflects the physical necessity of nonnegative abundances; FCLS, on the other hand, imposes additionally a sum-to-one constraint to further embody realistic conditions into the model.

We note that, recently, there have been methods proposed that deploy compressed sensing (CS) for dimensionality reduction, endmember extraction, and reconstruction of data. Such CS-based methods are attractive due to computational efficiency and ease of implementation [35, 36, 63, 67, 97]; however, most literature on CS-based methods is focused exclusively on achieving spectral dimensionality reduction such that a minimal reconstruction error (usually in the ℓ_1 norm) is achieved. We note, on the other hand, that it is not imperative to incur the computational burden associated with ℓ_1-norm reconstruction if our objective is classification in the reduced space rather than reconstruction [29, 30, 33, 66].

In this chapter, we address the significance of dimensionality reduction for the spectral-unmixing problem associated with hyperspectral imagery, coupling the random-projection methodology introduced in the previous chapters with the NNLS [12, 16, 56, 68, 78] strategy for spectral unmixing. In particular, we extend the concept of dimensionality reduction using HM-based projections to the hyperspectral-unmixing paradigm.

In general, it is expected that dimensionality reduction impacts spectral unmixing in ways such as increasing discrimination capability and abundance estimation through the decorrelation of spectral bands, through decreasing signal disparity, and through providing

inherent noise reduction while preserving essential information present in the data. While it would seem that a fully-constrained model that incorporates both nonnegativity and a sum-to-one condition provides the most accurate reflection of reality, recent work (e.g., [91]) suggests that merely imposing nonnegativity alone is sufficient. Consequently, in this work, we adopt only a nonnegativity assumption as it results in a dramatically simplified mathematical formulation.

Furthermore, it is important to note that random projections alone do not provide data-specific information discernment which can be crucial to spectral unmixing, wherein the goal is constituent endmember-abundance estimation. This task instead calls for a more detailed learning of data structure such as that provided by transform-based dimensionality-reduction methods. However, when applied directly to hyperspectral data, transform-based methods can be computationally heavy. Therefore, data-learning (in this case feature selection) can be performed in a random-projection domain to effectively exploit the benefits of a data-dependent representation while at the same time alleviating the computational burden of transform-based-methods. Hence, we employ HM-based random projections with feature selection using FSVD as described in Chapter 3. This is then followed by NNLS for abundance estimation to provide a structured two-stage dimensionality reduction tailored to the spectral-unmixing problem.

The rest of this chapter is organized as follows. In Section 4.2, we briefly overview the LMM as well as NNLS for spectral unmixing, while, in Section 4.3, we discuss our proposed combination of random projections, FSVD, and NNLS. In Section 4.4, we present

a battery of experimental results, while we make some concluding observations in Section 4.5. We note that the work presented in this chapter was initially published as [76].

4.2 NNLS and Abundance Estimation

In the LMM, a pixel vector with N spectral bands, $\mathbf{y} \in \mathbb{R}^N$, is described as a linear combination of C endmembers $\mathbf{M} \in \mathbb{R}^{N \times C}$ and their corresponding proportional abundances $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \dots & \alpha_C \end{bmatrix}^T$ such that

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{w}, \tag{4.1}$$

where $\mathbf{w} \in \mathbb{R}^N$ is an inherent noise process such as atmospheric turbulence, noise during signal acquisition, etc. In practice, we are given simply the vector $\mathbf{y}$ along with mixing matrix $\mathbf{M}$ with C endmembers that have been extracted from the data. Our aim is thus to estimate the abundance of the endmembers. In the NNLS formulation of this unmixing problem, we estimate the true abundance $\boldsymbol{\alpha}$ via a quadratically constrained optimization problem.

NNLS belongs to the set of constrained least-squares regression problems wherein variables are limited to nonnegative values. A typical NNLS problem can be defined as, given an input matrix $\mathbf{M} \in \mathbb{R}^{N \times C}$ and a pixel vector $\mathbf{y} \in \mathbb{R}^N$, find a nonnegative vector $\hat{\boldsymbol{\alpha}}(\mathbf{y}) \in \mathbb{R}^C$ such that

$$\hat{\boldsymbol{\alpha}}(\mathbf{y}) = \arg \min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^C \\ \boldsymbol{\alpha} \geq \mathbf{0}}} \left\| \mathbf{M}\boldsymbol{\alpha} - \mathbf{y} \right\|_2^2. \tag{4.2}$$

Nonnegativity is relevant with regard to spectral unmixing in hyperspectral imagery since we are estimating fractional abundance values of endmembers in the mixed pixel, and, physically, the assumption of nonnegativity always holds true.

As for classification, we obtain classification maps from the estimated abundance values by hard classification, i.e.,

$$\theta(\mathbf{y}) = \arg \max_{c \in \{1,\ldots,C\}} \hat{\alpha}_c(\mathbf{y}), \tag{4.3}$$

where $\hat{\boldsymbol{\alpha}}(\mathbf{y}) = \begin{bmatrix} \hat{\alpha}_1(\mathbf{y}) & \ldots & \hat{\alpha}_C(\mathbf{y}) \end{bmatrix}^T$, and $\theta(\mathbf{y})$ is the class label assigned to vector $\mathbf{y}$ based on the class with maximum abundance. The reconstructed vector $\hat{\mathbf{y}}$ is then

$$\hat{\mathbf{y}} = \mathbf{M}\hat{\boldsymbol{\alpha}}(\mathbf{y}). \tag{4.4}$$

Below, we use (4.4) to formulate a reconstruction error with the goal of evaluating classification performance.

4.3 Proposed Approach

4.3.1 NNLS Based on Random Projections

Our main contribution here is to couple dimensionality reduction driven by random projections with the NNLS paradigm for spectral unmixing such that the unmixing is effectively applied in a reduced-dimensional space, thereby ameliorating computational aspects of the problem. To perform dimensionality reduction using random projections, we use either the GM-based projection matrix $\mathbf{P}_{\text{GM}}$ from (2.1) or the HM-based projection matrix $\mathbf{P}_{\text{HM}}$ from (2.3). Assuming that the endmember matrix $\mathbf{M}$ is known, its corresponding low-dimensional representation using HM-based projections is

$$\hat{\mathbf{M}}_{\text{HM}} = \mathbf{P}_{\text{HM}}^T \mathbf{M} \in \mathbb{R}^{K \times C}. \tag{4.5}$$

By combining equations (4.2) with (2.3) and (4.5), we couple the Hadamard-based dimensionality reduction to the NNLS problem (i.e., HM-NNLS) as

$$\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}) = \arg\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^C \\ \boldsymbol{\alpha}\geq\mathbf{0}}} \left\|\hat{\mathbf{M}}_{\text{HM}}\boldsymbol{\alpha} - \mathbf{P}_{\text{HM}}^T\mathbf{y}\right\|_2^2 \tag{4.6}$$

which applies the NNLS formulation of (4.2) with dimensionality reduction into a K-dimensional space, $K \ll N$. As noted in Section 2.2.2, random projection driven by a Hadamard matrix is computationally efficient because it is implementable as a series of addition and subtraction operations.

As an alternative to HM-NNLS, we can instead use a Gaussian-based projection matrix, $\mathbf{P}_{\text{GM}}$, The resulting Gaussian-based dimensionality reduction with NNLS (i.e., GM-NNLS) is then obtained by substituting $\mathbf{P}_{\text{GM}}$ into (4.5) and (4.6):

$$\hat{\mathbf{M}}_{\text{GM}} = \mathbf{P}_{\text{GM}}^T\mathbf{M} \in \mathbb{R}^{K\times C}, \tag{4.7}$$

$$\hat{\boldsymbol{\alpha}}_{\text{GM}}(\mathbf{y}) = \arg\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^C \\ \boldsymbol{\alpha}\geq\mathbf{0}}} \left\|\hat{\mathbf{M}}_{\text{GM}}\boldsymbol{\alpha} - \mathbf{P}_{\text{GM}}^T\mathbf{y}\right\|_2^2 \tag{4.8}$$

The drawback to GM-NNLS is a heavier computational load due to the dense matrix-multiplication computations implied by the Gaussian matrix. The resulting HM-NNLS and GM-NNLS algorithms are detailed in Algorithm 4.1, where the corresponding GM-NNLS algorithm follows by substituting $\mathbf{P}_{\text{GM}}$ for $\mathbf{P}_{\text{HM}}$.

4.3.2 Random Projection with Feature Selection Using FSVD-Based NNLS

Although dimensionality reduction using random projections achieves reduction in time and space complexity, the spectral-unmixing problem might benefit from learning the underlying data structure in order to achieve better abundance estimation and classification

performance. For this, it would be possible to apply any data-dependent transform; however, the FSVD paradigm proposed in Chapter 3, which combines the benefits of both computationally lightweight random projections and a data-dependent transform-based representation offers an attractive alternative. We formulate the resulting HM-FSVD-NNLS and GM-FSVD-NNLS as follows.

Let the collection of target vectors be defined as $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \dots \mathbf{y}_M \end{bmatrix} \in \mathbb{R}^{N \times M}$. This is randomly projected using HM-based projections to an R-dimensional space. SVD is then performed on the resulting lower-dimensional $\hat{\mathbf{Y}}_{\text{HM}}$, and the desired K features are selected for further dimensionality reduction such that $K < R \ll N$. The selected K features are used as the new projection matrix $\hat{\mathbf{P}}_{\text{HM}}$ for a second-stage dimensionality reduction. This new projection matrix is then used in (4.5) and (4.6) to drive the projection-domain NNLS. The proposed HM-FSVD-NNLS and GM-FSVD-NNLS algorithms are detailed in Algorithm 4.2, where the corresponding GM-FSVD-NNLS algorithm follows by substituting $\mathbf{P}_{\text{GM}}$ for $\mathbf{P}_{\text{HM}}$.

4.4 Experimental Results

4.4.1 Experimental Setup

In this section, we experimentally validate the efficacy of our proposed random-projection-based dimensionality reduction and abundance estimation. We use the same Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) dataset that was used in Section 3.3.2, i.e., Salinas-A. We also generate an artificial random dataset with $N = 204$ spectral bands and $M = 2,000$ samples by linearly mixing pixels from Salinas-A. The

number of endmembers is $C = 6$; these are collected in mixing matrix $\mathbf{M}$. Artificially generated abundance vector $\boldsymbol{\alpha}^{(m)}$ for pixel m is modeled as a uniformly distributed random variable within $[0, 1]$. Finally, we add additive white Gaussian noise $\eta^{(m)}$ such that the final artificial pixel vector for pixel m is

$$\mathbf{y}^{(m)} = \mathbf{M}\boldsymbol{\alpha}^{(m)} + \eta^{(m)}. \tag{4.9}$$

Finally, we use the Cuprite[1] dataset, an AVIRIS image acquired over the Cuprite areas in Nevada. This dataset has a spatial dimension of 512×614 with 224 spectral bands; after removal of water absorption bands (bands 1–2, 105–115, 150–170, and 223–224), the remaining 188 spectral bands are used. Figure 4.11 gives a grayscale visualization of Cuprite.

For quantitative assessment, we define the average abundance error (AE) over the entire M-pixel dataset as

$$\text{AE} = \frac{1}{M}\sum_{m=1}^{M} \left\|\boldsymbol{\alpha}^{(m)} - \hat{\boldsymbol{\alpha}}(\mathbf{y}^{(m)})\right\|^2, \tag{4.10}$$

where $\hat{\boldsymbol{\alpha}}(\mathbf{y}^{(m)})$ is the NNLS-estimate (i.e., (4.2) or (4.6), depending on whether random projection is used or not) for pixel vector $\mathbf{y}^{(m)}$. We also define an average pixel reconstruction error (PRE) as the error between original target vector $\mathbf{y}^{(m)}$ and the reconstructed target vector $\hat{\mathbf{y}}^{(m)}$ as

$$\text{PRE} = \frac{1}{M}\sum_{m=1}^{M} \left\|\mathbf{y}^{(m)} - \hat{\mathbf{y}}^{(m)}\right\|^2, \tag{4.11}$$

where $\hat{\mathbf{y}}^{(m)}$ is either $\mathbf{M}\hat{\boldsymbol{\alpha}}(\mathbf{y})$, $\hat{\mathbf{M}}_{\text{HM}}\hat{\boldsymbol{\alpha}}(\mathbf{y})$, or $\hat{\mathbf{M}}_{\text{GM}}\hat{\boldsymbol{\alpha}}(\mathbf{y})$ as appropriate, depending on whether random projections are used and which kind.

[1] `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`

In the experimental results to follow, we compare the methods proposed in this chapter, namely, HM-NNLS and GM-NNLS as described in Section 4.3.1 as well as HM-FSVD-NNLS and GM-FSVD-NNLS as described in Section 4.3.2. Additionally, we compare to dimensionality reduction using SVD applied directly to the original dataset followed by NNLS, which we denote as "SVD-NNLS," and, finally, NNLS abundance estimation applied directly on the original dataset without dimensionality reduction (denoted simply as "NNLS").

For HM-FSVD-NNLS and GM-FSVD-NNLS, a two-stage dimensionality reduction is performed by using random projections to first reduce to R dimensions followed by a second reduction to a final dimensionality K; as in Section 3.3, we use $K = 0.7R$ throughout the results presented here. On the other hand, in the case of HM-NNLS, GM-NNLS, and SVD-NNLS, dimensionality reduction to K dimensions was carried out directly. We use specifically $K \in \{14, 29, 43, 57, 71, 85, 100, 114, 129, 143\}$.

4.4.2 Results for the Artificial Dataset

Tables 4.1 and 4.2 present the AE and PRE, respectively, for the artificial dataset, using a reduced dimensionality of $K = 29$ for the HM-NNLS and GM-NNLS techniques. We see that the methods based on random projections provide close approximation to the original data with low AE and PRE regardless of the noise level; this effect can be especially noted in the case of HM-FSVD-NNLS which uniformly achieves the lowest PRE of all the methods considered.

4.4.3 Results for the Salinas-A Dataset

For the Salinas-A dataset, Figure 4.3 illustrates the spectral angle (in degrees) between the eigenvectors produced by the FSVD-NNLS methods and those from SVD as determined directly from the original dataset with full dimensionality. We see that HM-FSVD-NNLS yields lower spectral angles than GM-FSVD-NNLS, proving that FSVD-methods does indeed provide effective data-learning and preserves essential information present in the data. Tables 4.3–4.5 tabulate overall classification accuracy (OA) and PRE for the Salinas-A dataset. Figures 4.4–4.9 give the fractional abundance maps for $K = 29$ for all methods under consideration, while Figure 4.10 illustrates the classification maps after hard classification (i.e., (4.3)) is performed on estimated abundance maps for a reduced dimension of $K = 29$. Finally, Table 4.6 gives the computation time for the various methods. We note that the methods based on random projections have computation times lower than both NNLS and SVD on the original dataset, and we see that HM-FSVD-NNLS generally provides the closest approximation to the original data while preserving important information so as to yield generally the lowest PRE and highest OA through all dimensions.

4.4.4 Results for the Cuprite Dataset

Figures 4.12–4.17 illustrate the estimated fractional abundance maps for reduced dimension $K = 92$ for the Cuprite dataset for all methods under consideration. We note that, visually, it appears that HM-FSVD-NNLS yields a more accurate abundance estimation compared to the other methods. Table 4.7 tabulates computation time for the

Cuprite dataset, again demonstrating that the random-projection-based methods are faster than SVD on the original data space.

4.5 Observations

In this chapter, we empirically demonstrated the effectiveness of using various methods for dimensionality reduction driven by random projections coupled with NNLS in order to address the spectral-unmixing problem in a reduced-dimensional space, thereby circumventing high storage and computation costs. The proposed HM-FSVD-NNLS incorporates benefits of both computationally lightweight random projections as well as the data-specific learning of transform-based methods, but with a significant reduction in computation time over SVD as applied in the original data space. All the methods based on random projections performed competitively; in particular, HM-FSVD-NNLS had the best abundance estimation and classification performance with low reconstruction errors even at low dimensions. As expected, all the HM-based methods were more computationally efficient than their GM-based counterparts, and significantly faster than SVD applied alone to the original dataset.

These observations conclude our investigations of random projections for dimensionality reduction of hyperspectral imagery. The next—and final—chapter of this dissertation makes a number of concluding remarks.

Table 4.1

AE for the artificial dataset over different noise levels for a reduced dimensionality of $K = 29$

Noise (in dB)	AE					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
20	**0.116**	**0.116**	0.153	0.137	0.190	0.121
40	0.058	0.061	0.076	0.077	**0.046**	0.062
60	0.020	0.025	0.052	0.024	**0.011**	0.021
80	3.72e-04	5.15e-04	4.55e-03	5.60e-04	**2.29e-04**	3.89e-04
100	3.73e-06	5.06e-06	5.22e-05	5.19e-06	**3.34e-06**	3.90e-06

Table 4.2

PRE for the artificial dataset over different noise levels for a reduced dimensionality of $K = 29$

Noise (in dB)	PRE					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
20	0.729	**0.702**	0.853	0.817	0.749	0.768
40	0.726	**0.694**	0.748	0.804	0.730	0.755
60	0.725	**0.693**	0.746	0.804	0.721	0.755
80	0.725	**0.693**	0.747	0.801	0.721	0.754
100	0.725	**0.693**	0.747	0.801	0.721	0.754

Table 4.3

OA and κ statistics for Salinas-A with reduced dimensionality $K = 29$

Class	**HM-NNLS**	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
1	**99.74**	**99.74**	**99.74**	**99.74**	**99.74**	**99.74**
2	38.49	**43.04**	38.49	39.24	38.37	38.49
3	96.49	96.62	96.33	96.59	92.34	**96.91**
4	98.68	**98.70**	98.61	98.61	96.13	98.68
5	**99.85**	**99.85**	**99.85**	**99.85**	**99.85**	**99.85**
6	99.25	**99.87**	99.15	99.37	98.99	99.49
OA	83.68	**84.33**	83.65	83.81	83.22	83.71
κ	80.07	**80.99**	80.02	80.15	79.76	80.12

Table 4.4

OA over varying reduced dimensionality K for Salinas-A

K	OA					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
14	83.71	**84.46**	83.64	83.02	82.81	83.71
29	83.68	**84.33**	83.65	83.81	83.22	83.71
43	83.68	83.72	83.65	**83.78**	83.39	83.71
57	83.68	83.55	83.65	83.02	83.23	**83.71**
71	83.68	**84.14**	83.67	82.66	83.58	83.71
85	83.68	**83.90**	83.67	83.43	83.24	83.71
100	83.71	**83.94**	83.71	83.43	83.67	83.71
114	83.71	**84.74**	83.71	83.73	83.71	83.71
129	83.71	**83.90**	83.71	83.76	83.60	83.71
143	83.71	**84.11**	83.71	83.71	83.52	83.71

Table 4.5

PRE over varying reduced dimensionality K for Salinas-A

K	PRE					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
14	0.0032	**0.0020**	0.0048	0.0035	0.0035	0.0030
29	0.0028	0.0026	**0.0025**	0.0033	0.0037	0.0030
43	0.0028	**0.0021**	0.0034	0.0033	0.0033	0.0030
57	0.0026	**0.0022**	0.0027	0.0031	0.0035	0.0030
71	0.0027	0.0024	**0.0023**	0.0032	0.0038	0.0030
85	0.0028	**0.0020**	0.0029	0.0032	0.0037	0.0030
100	**0.0024**	0.0026	**0.0024**	0.0031	0.0031	0.0030
114	0.0025	**0.002**1	0.0030	0.0031	0.0031	0.0030
129	0.0028	**0.0022**	0.0033	0.0030	0.0032	0.0030
143	0.0030	**0.0024**	0.0035	0.0030	0.0032	0.0030

Table 4.6

Computation time (in sec) over varying reduced dimensionality K for Salinas-A

K	Computation time (in sec)					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
14	**2.19**	2.45	2.59	3.65	**4.72**	4.20
29	**2.21**	2.54	2.75	3.54	**4.59**	4.20
43	**2.23**	2.41	2.64	4.01	**4.31**	4.20
57	**2.52**	2.77	3.03	3.39	**4.33**	4.20
71	**2.47**	3.09	2.82	3.61	**4.32**	4.20
85	**2.55**	2.81	2.71	3.56	**4.82**	4.20
100	**2.58**	3.18	2.86	3.57	**4.89**	4.20
114	**2.64**	2.81	2.94	3.77	**5.19**	4.20
129	**2.70**	2.92	3.48	3.51	**5.29**	4.20
143	**2.63**	2.81	3.46	3.54	**5.34**	4.20

Table 4.7

Computation time (in sec) for reduced dimensionality $K = 92$ for Cuprite

K	Computation time (in sec)					
	HM-NNLS	**HM-FSVD-NNLS**	**GM-NNLS**	**GM-FSVD-NNLS**	**SVD-NNLS**	**NNLS**
92	301	412	328	449	3857	1031

1. **Input:** Endmember matrix $\mathbf{M} \in \mathbb{R}^{N \times C}$ and a test hyperspectral vector $\mathbf{y} \in \mathbb{R}^N$.
2. Generate random projection matrix $\mathbf{P}_{\text{HM}}$ via (2.3).
3. Use computed random projections to project endmember matrix $\mathbf{M}$ to its corresponding K-dimensional space producing $\hat{\mathbf{M}}_{\text{HM}}$:
$$\hat{\mathbf{M}}_{\text{HM}} = \mathbf{P}_{\text{HM}}^T \mathbf{M} \in \mathbb{R}^{K \times C}.$$
4. Conduct the NNLS unmixing problem for abundance estimation:
$$\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}) = \arg\min_{\boldsymbol{\alpha} \in \mathbb{R}^C \, \boldsymbol{\alpha} \geq \mathbf{0}} \left\| \hat{\mathbf{M}}_{\text{HM}} \boldsymbol{\alpha} - \mathbf{P}_{\text{HM}}^T \mathbf{y} \right\|_2^2.$$
5. Perform hard classification on the estimated abundance values to derive class labels,
$$\theta(\mathbf{y}) = \arg\max_{c \in \{1,\ldots,C\}} \hat{\alpha}_c(\mathbf{y}),$$
where $\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}) = \begin{bmatrix} \hat{\alpha}_1(\mathbf{y}) & \cdots & \hat{\alpha}_C(\mathbf{y}) \end{bmatrix}$.
6. Compute the reconstructed target vector $\hat{\mathbf{y}}$ from the estimated abundances:
$$\hat{\mathbf{y}}_{\text{HM}} = \hat{\mathbf{M}} \hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}).$$
7. **Output:** Abundances $\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y})$ and hard-classification result $\theta(\mathbf{y})$.

Figure 4.1

The randomized HM-NNLS algorithm for abundance estimation

1. **Input:** Endmember matrix $\mathbf{M} \in \mathbb{R}^{N\times C}$, original dataset $\mathbf{Y} \in \mathbb{R}^{N\times M}$, and a test hyperspectral vector $\mathbf{y} \in \mathbb{R}^N$.

2. Generate random projection matrix $\mathbf{P}_{\text{HM}}$ via (2.3).

3. Reduce dimensionality of dataset $\mathbf{Y}$:

$$\hat{\mathbf{Y}}_{\text{HM}} = \mathbf{P}_{\text{HM}}^T \mathbf{Y}.$$

4. Orthonormalize the rows of $\hat{\mathbf{Y}}_{\text{HM}}$ to produce $\mathbf{Q} = \text{orth}(\hat{\mathbf{Y}}_{\text{HM}}) \in \mathbb{R}^{R\times M}$

5. Perform SVD on $\mathbf{YQ}^T \in \mathbb{R}^{N\times R}$. i.e., $\mathbf{YQ}^T = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T$

6. Let the new projection matrix be $\hat{\mathbf{P}}_{\text{HM}} \in \mathbb{R}^{N\times K}$ contain the largest K singular vectors from $\hat{\mathbf{U}}$ such that $K < R \ll N$.

7. Let the new reduced-dimension mixing matrix be

$$\hat{\mathbf{M}}_{\text{HM}} = \hat{\mathbf{P}}_{\text{HM}}^T \mathbf{M} \in \mathbb{R}^{K\times C}.$$

8. Conduct the NNLS unmixing problem for abundance estimation:

$$\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}) = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^C \boldsymbol{\alpha}\geq \mathbf{0}} \left\|\hat{\mathbf{M}}_{\text{HM}}\boldsymbol{\alpha} - \hat{\mathbf{P}}_{\text{HM}}^T\mathbf{y}\right\|_2^2.$$

9. Perform hard classification on the estimated abundance values to derive class labels,

$$\theta(\mathbf{y}) = \arg\max_{c\in\{1,\ldots,C\}} \hat{\alpha}_c(\mathbf{y}),$$

where $\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}) = \begin{bmatrix}\hat{\alpha}_1(\mathbf{y}) & \cdots & \hat{\alpha}_C(\mathbf{y})\end{bmatrix}$.

10. Compute the reconstructed target vector $\hat{\mathbf{y}}$ from the estimated abundances:

$$\hat{\mathbf{y}}_{\text{HM}} = \hat{\mathbf{M}}\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y}).$$

11. **Output:** Abundances $\hat{\boldsymbol{\alpha}}_{\text{HM}}(\mathbf{y})$ and hard-classification result $\theta(\mathbf{y})$.

Figure 4.2

The randomized HM-FSVD-NNLS algorithm for abundance estimation

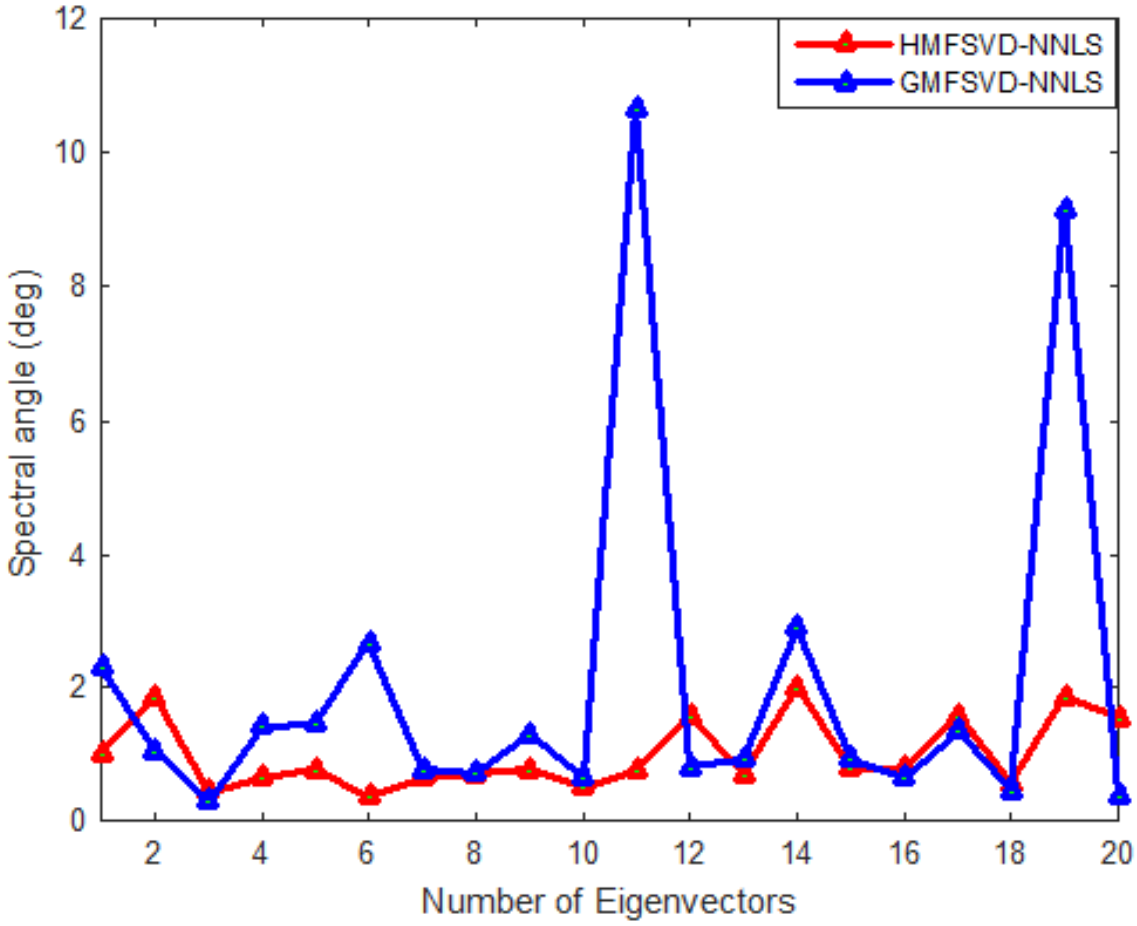
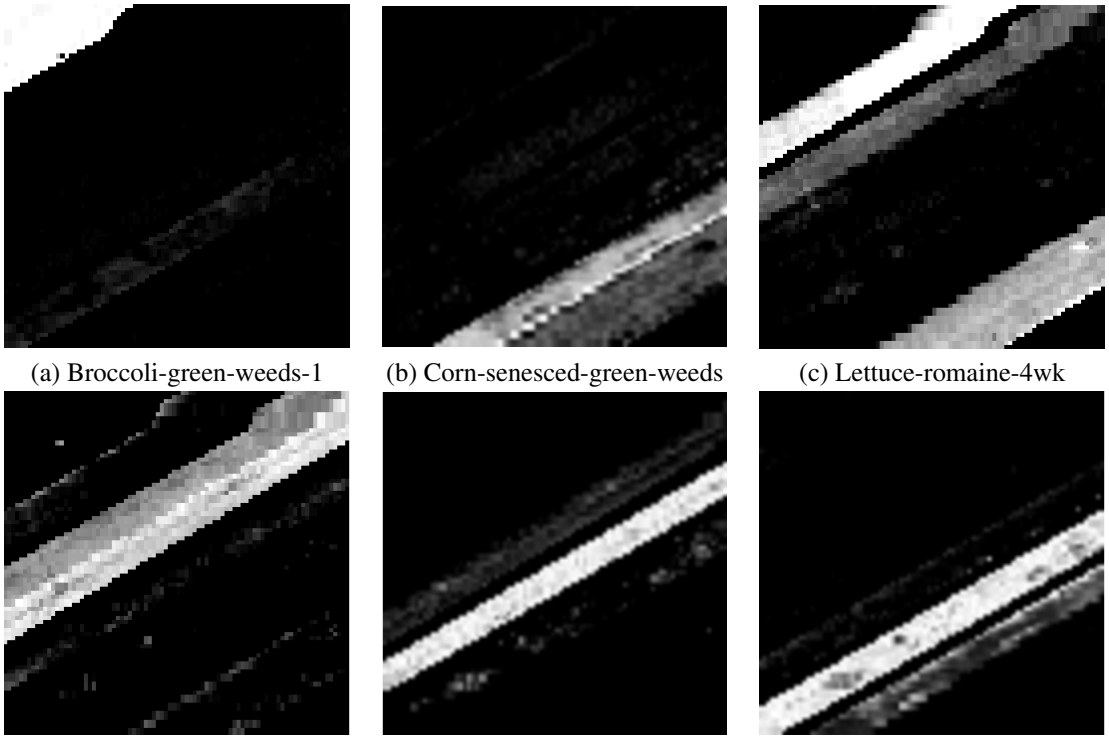


Figure 4.3

Spectral angle between the eignevectors produced by the FSVD-SVD-NNLS methods for Salinas-A

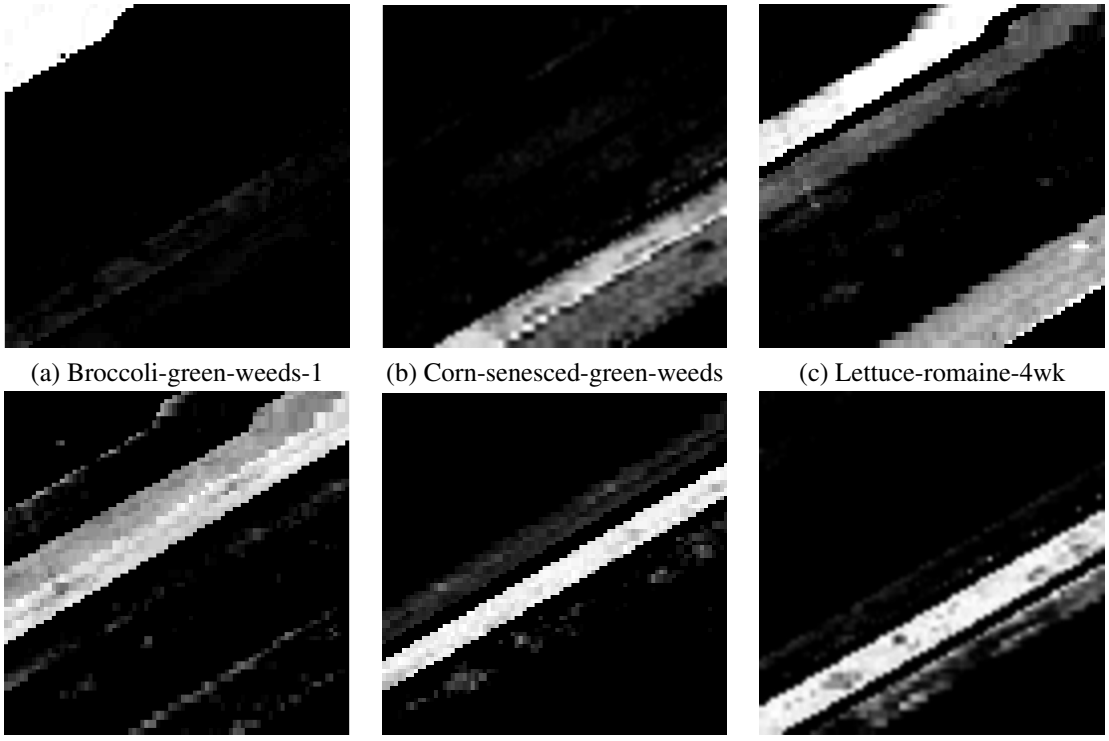


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.4

Abundance maps for each class using HM-NNLS for the Salinas-A dataset for $K = 29$

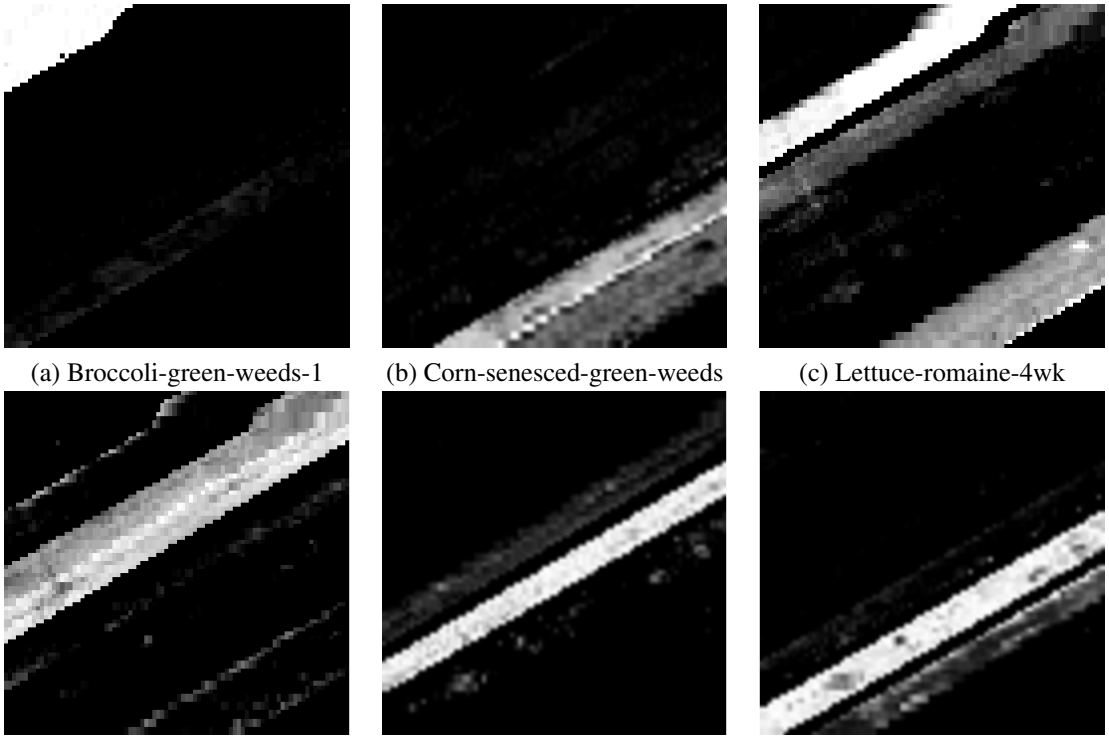


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.5

Abundance maps for each class using HM-FSVD-NNLS for the Salinas-A dataset for $K = 29$

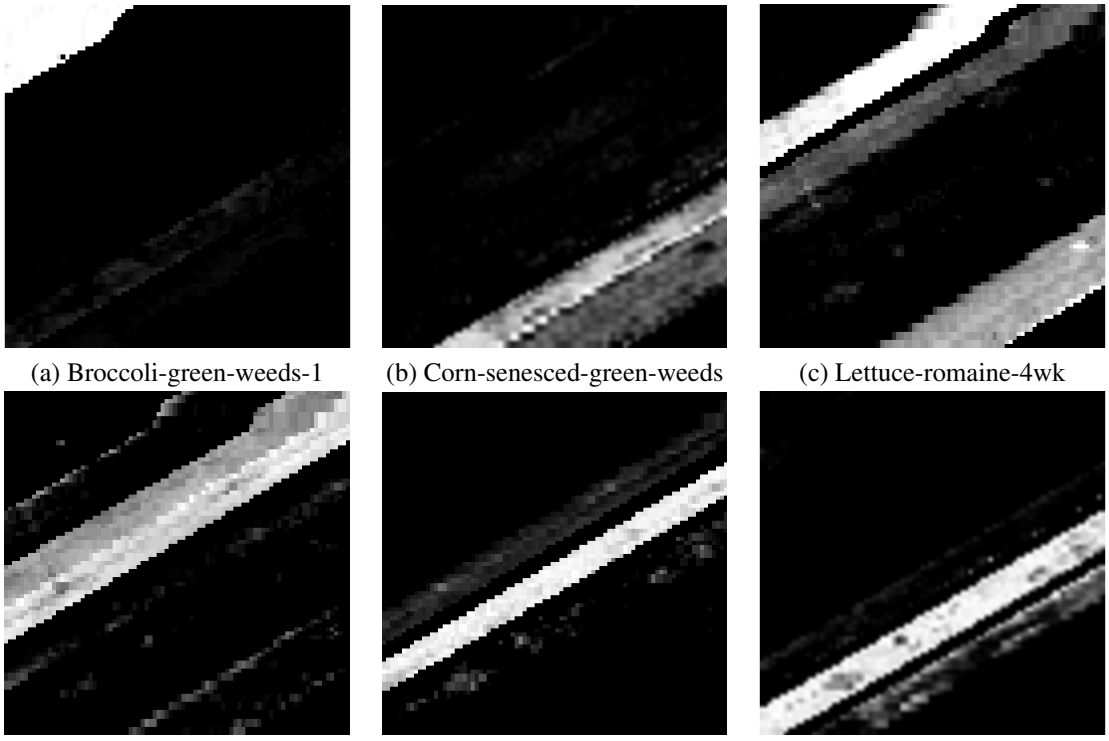


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.6

Abundance maps for each class using GM-NNLS for the Salinas-A dataset for $K = 29$

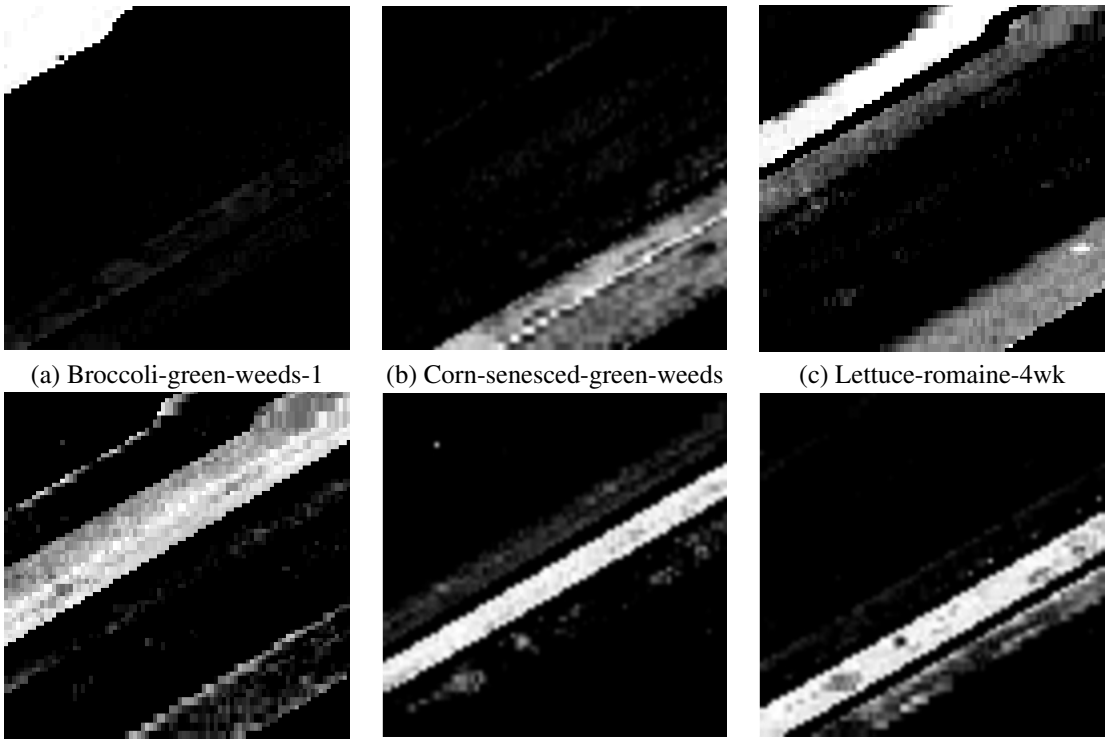


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.7

Abundance maps for each class using GM-FSVD-NNLS for the Salinas-A dataset for $K = 29$

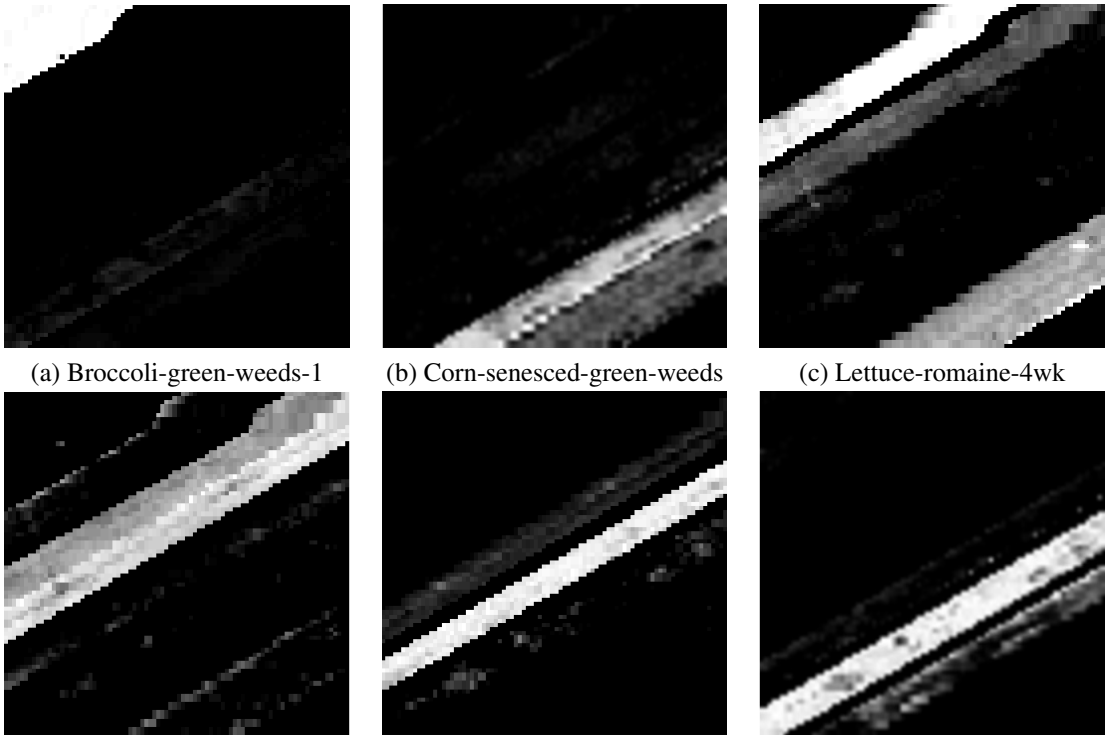


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.8

Abundance maps for each class using SVD-NNLS for the Salinas-A dataset for $K = 29$

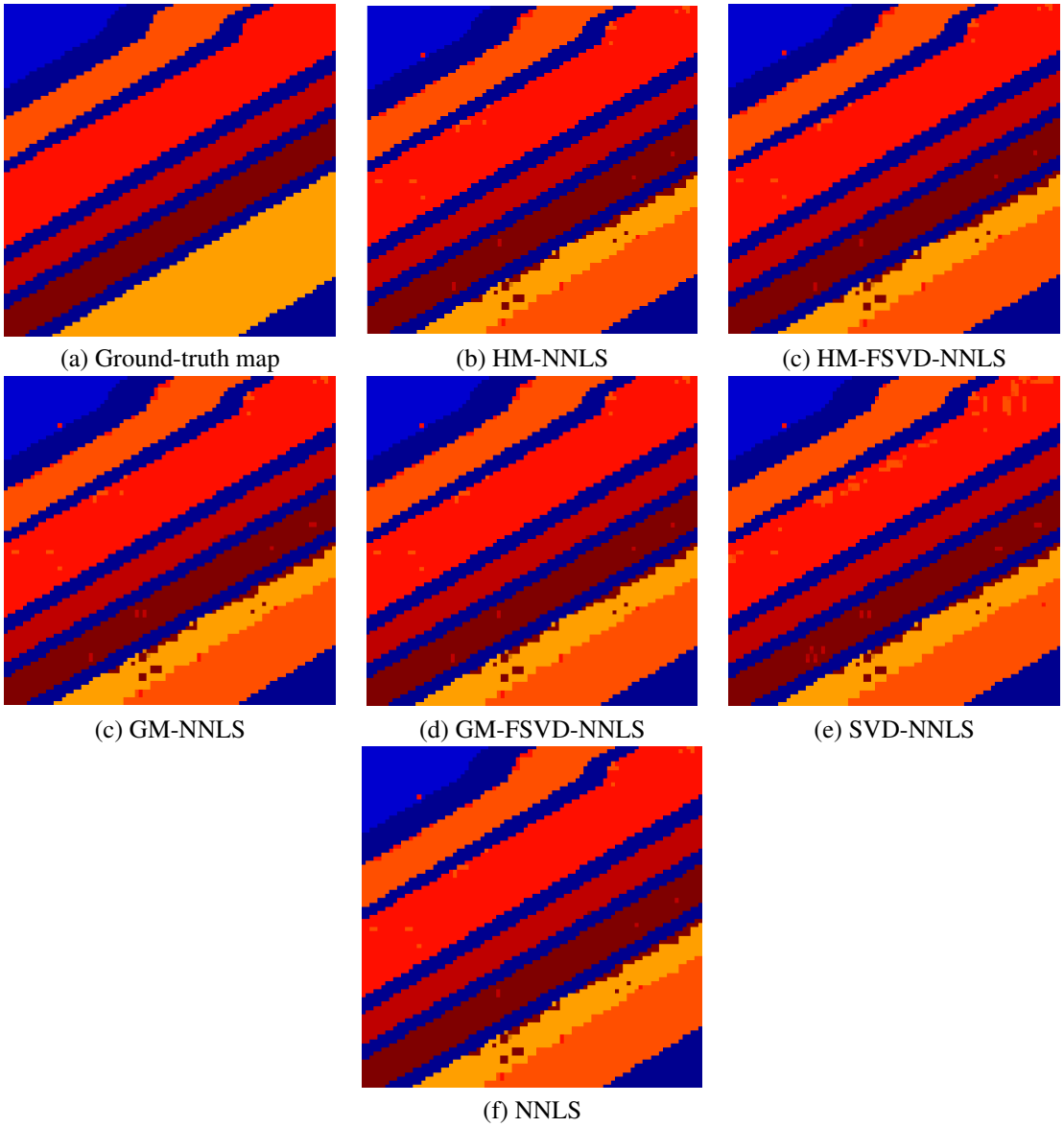


(a) Broccoli-green-weeds-1 (b) Corn-senesced-green-weeds (c) Lettuce-romaine-4wk

(c) Lettuce-romaine-5wk (d) Lettuce-romaine-6wk (e) Lettuce-romaine-7wk

Figure 4.9

Abundance maps for each class using NNLS for the Salinas-A dataset.



(a) Ground-truth map (b) HM-NNLS (c) HM-FSVD-NNLS

(c) GM-NNLS (d) GM-FSVD-NNLS (e) SVD-NNLS

(f) NNLS

Figure 4.10

Classification maps for Salinas-A dataset illustrating different methods for reduced dimension $K = 29$

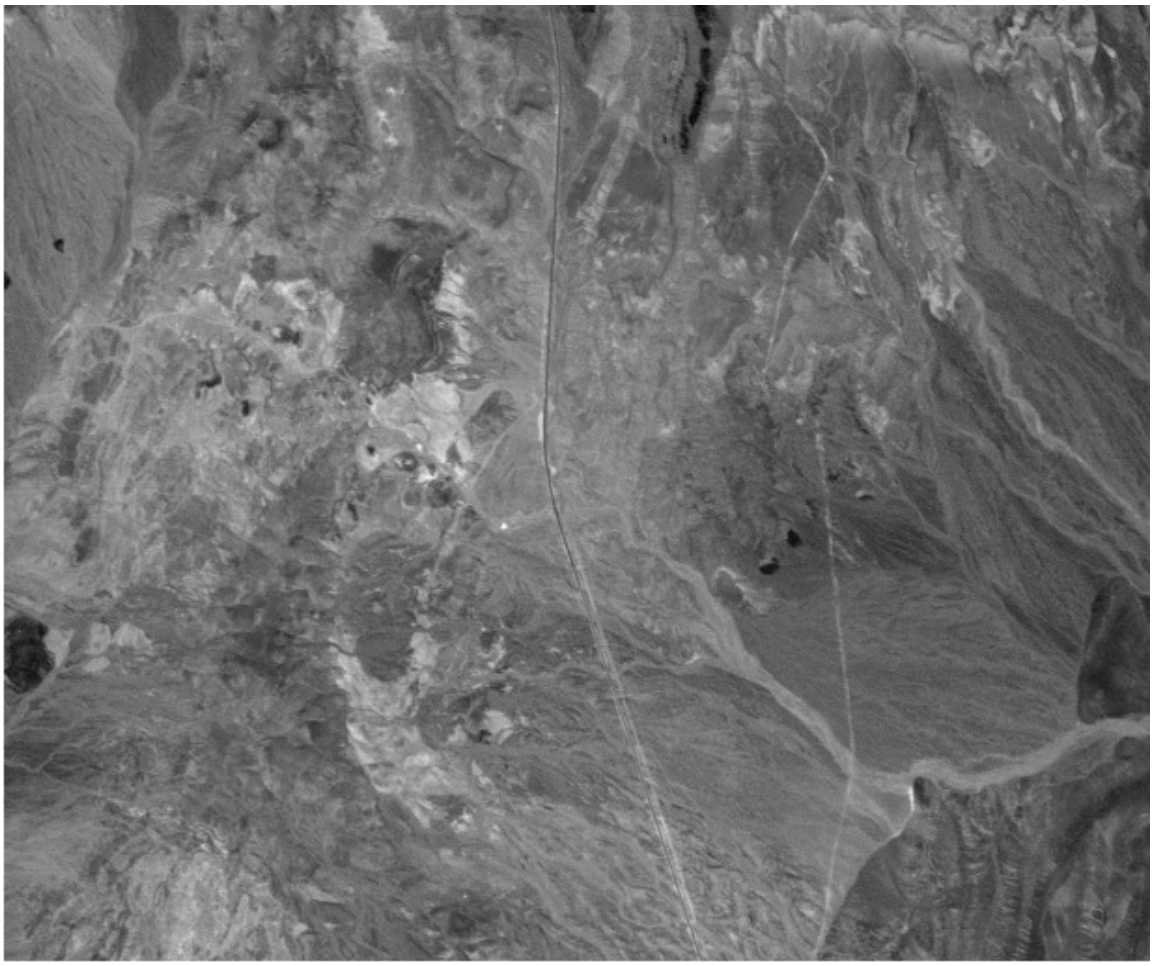
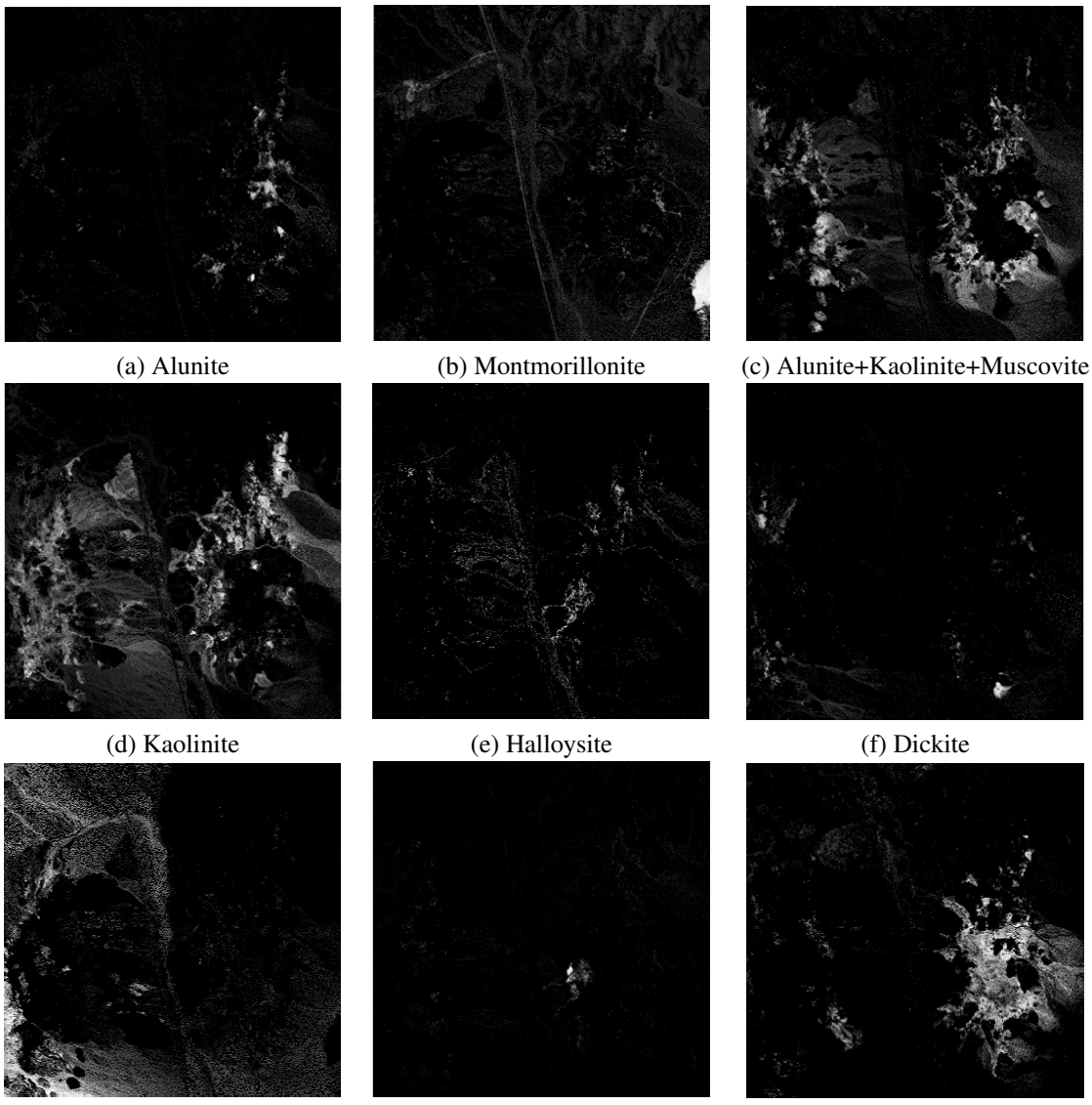


Figure 4.11

Grayscale visualization of the Cuprite dataset



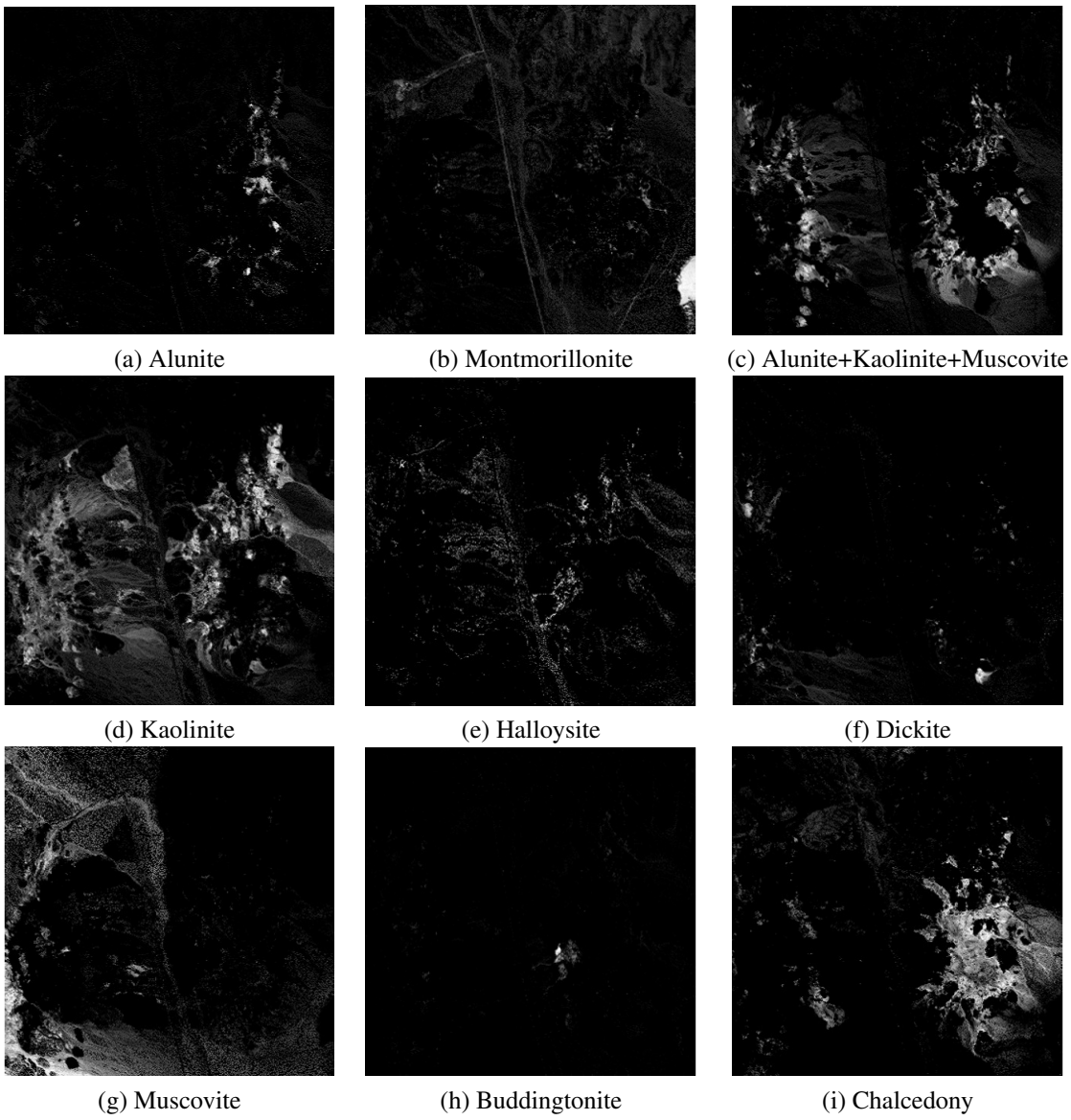
(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.12

Abundance maps for each endmember using HM-NNLS for the Cuprite dataset with $K = 92$



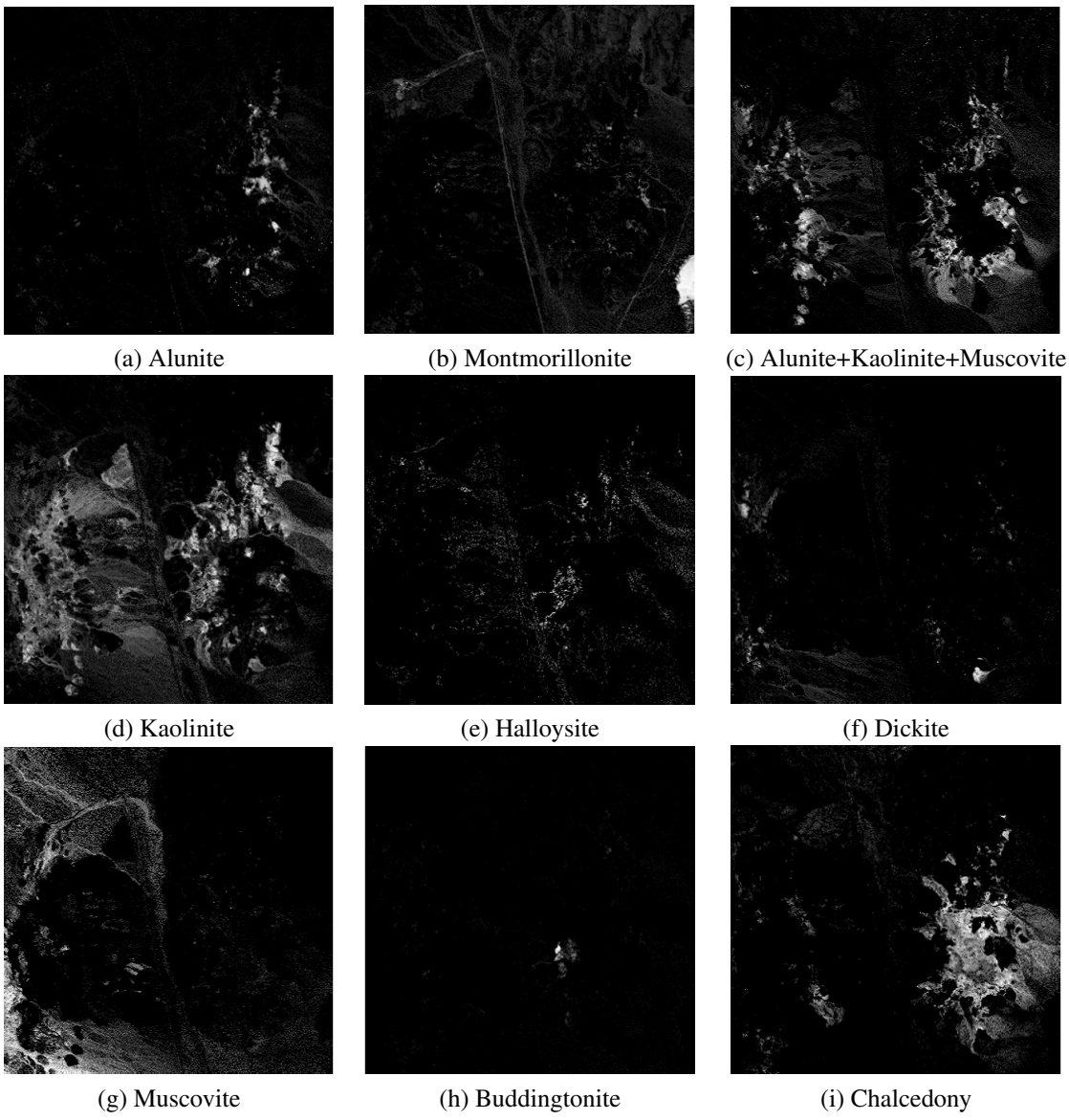
(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.13

Abundance maps for each endmember using HM-FSVD-NNLS for the Cuprite dataset with $K = 92$



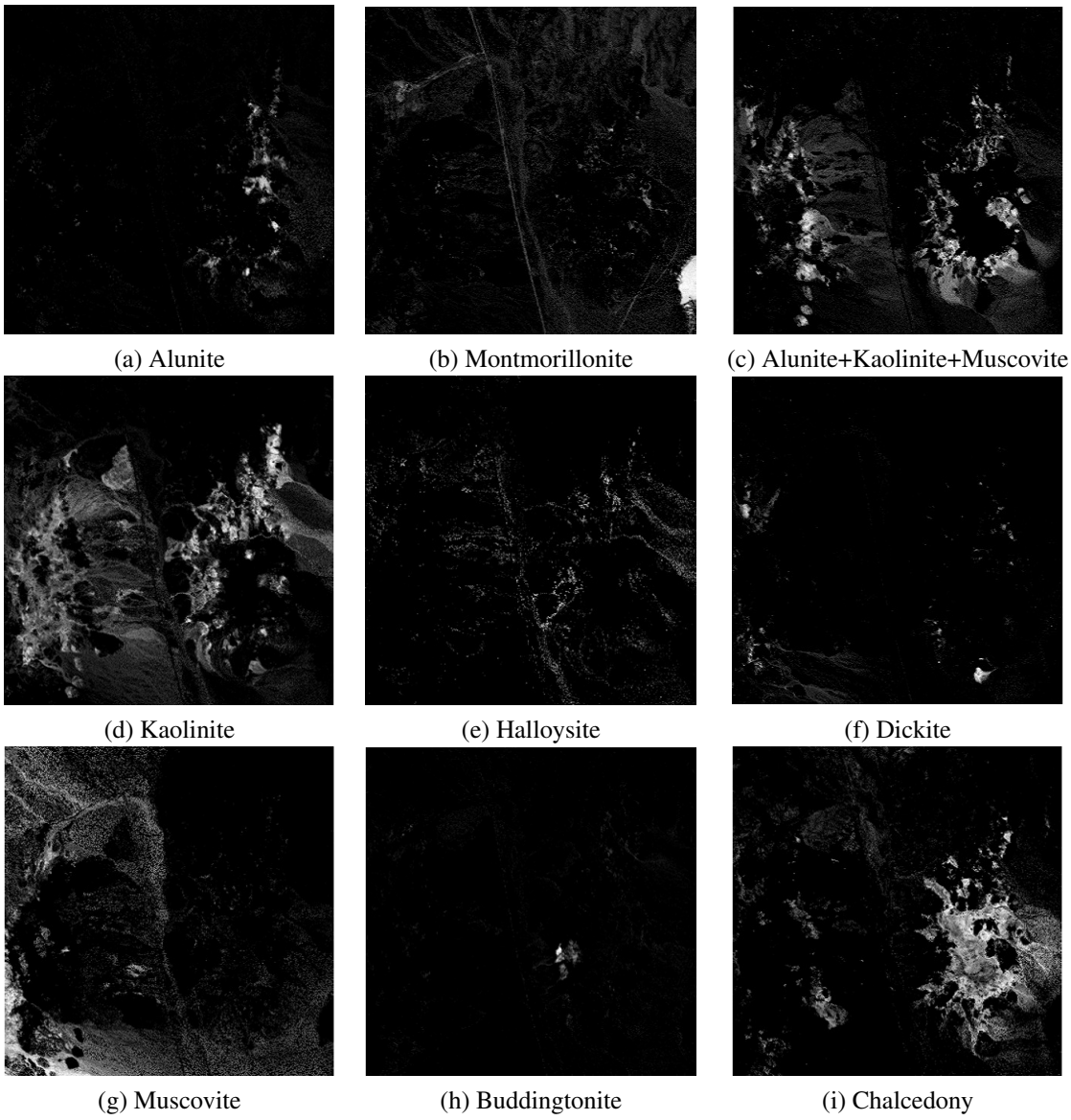
(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.14

Abundance maps for each endmember using GM-NNLS for the Cuprite dataset with $K = 92$



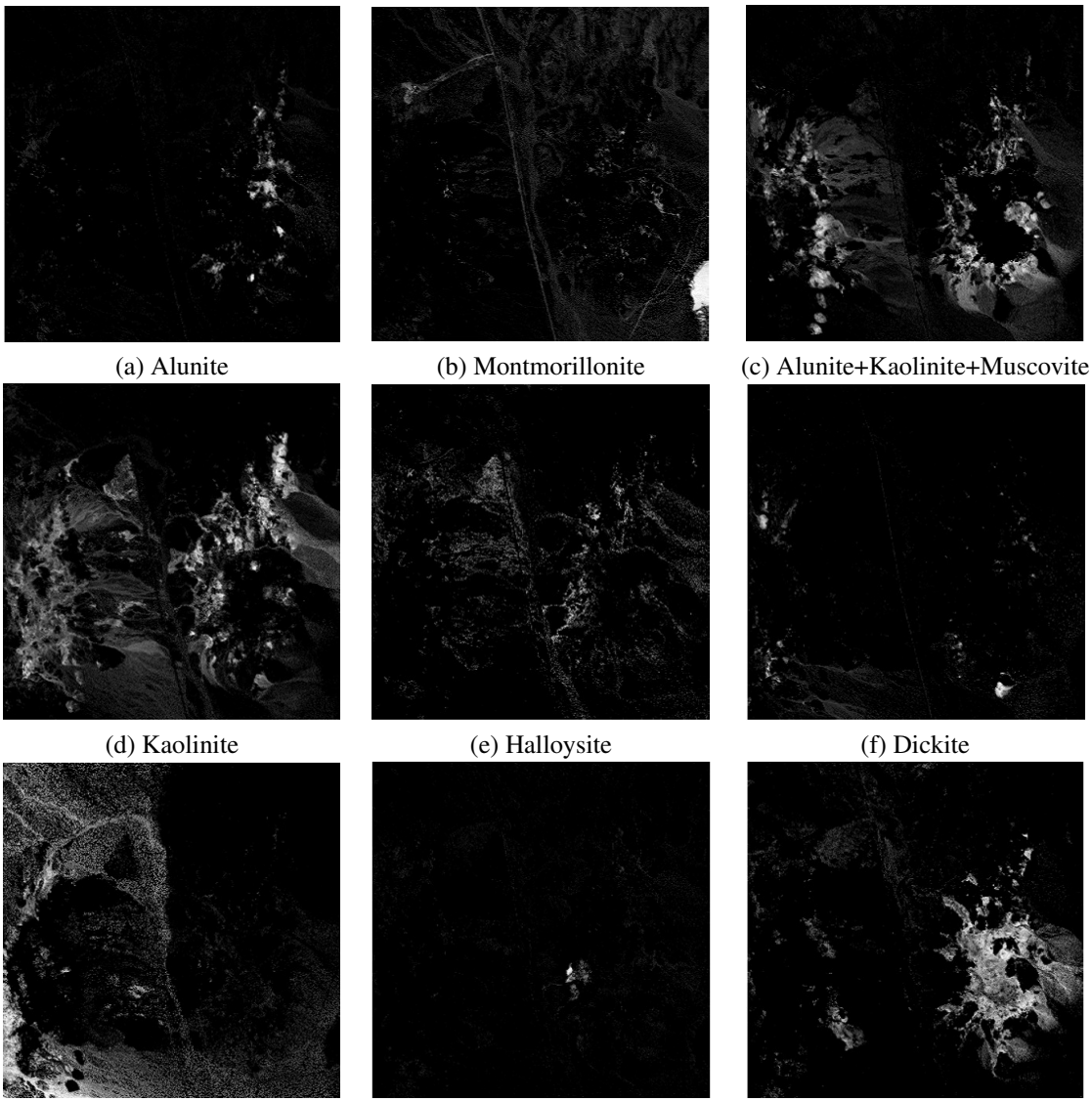
(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.15

Abundance maps for each endmember using GMFSVD-NNLS for the Cuprite dataset with $K = 92$



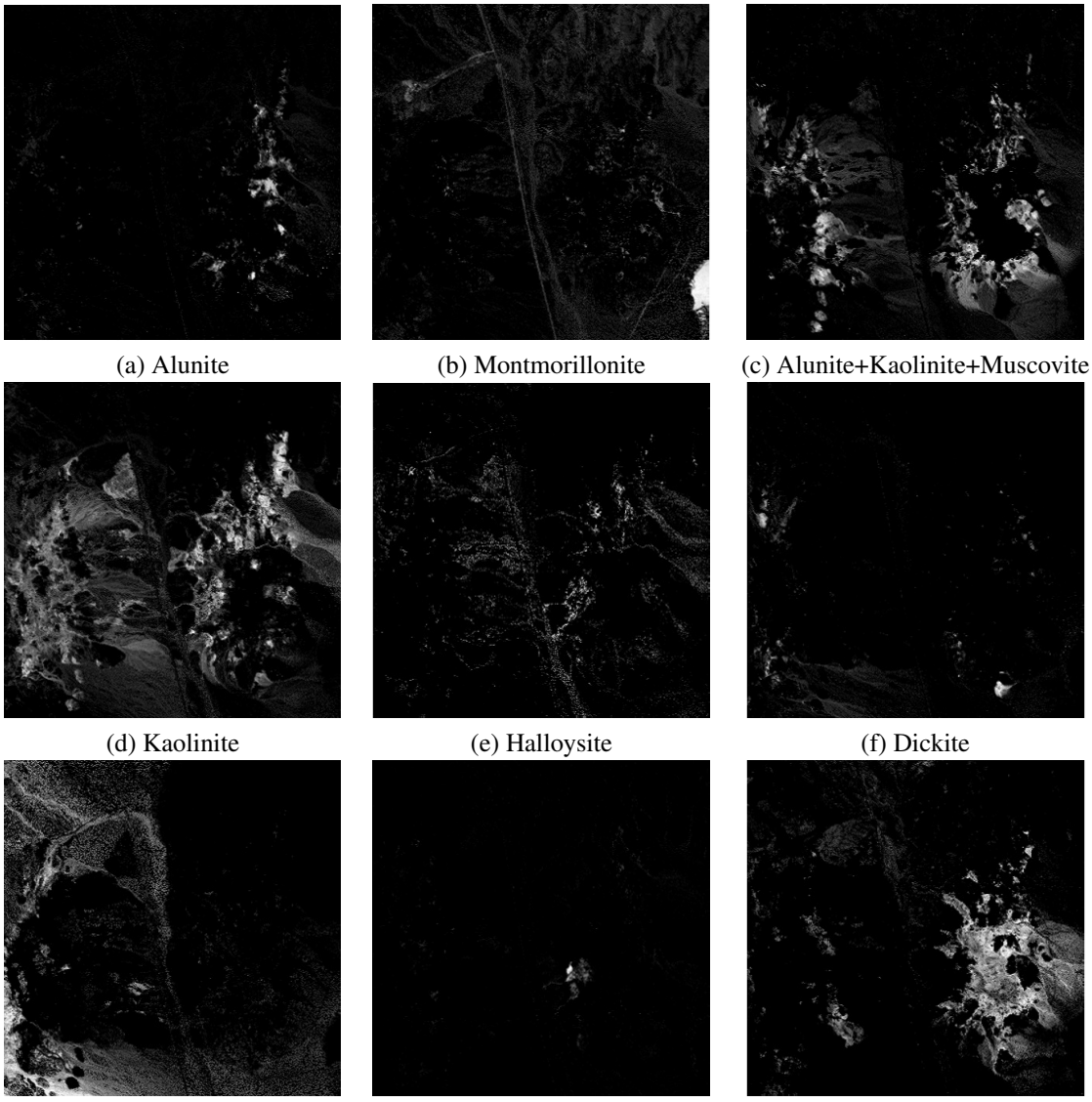
(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.16

Abundance maps for each endmember using SVD-NNLS for the Cuprite dataset with$K = 92$



(a) Alunite (b) Montmorillonite (c) Alunite+Kaolinite+Muscovite

(d) Kaolinite (e) Halloysite (f) Dickite

(g) Muscovite (h) Buddingtonite (i) Chalcedony

Figure 4.17

Abundance maps for each endmember using NNLS for the Cuprite dataset.

CHAPTER 5

CONCLUSIONS

This dissertation targets the development of algorithms for the dimensionality reduction of hyperspectral imagery using random projection. There has been a growing need for faster and more efficient dimensionality-reduction techniques—especially in the case of hyperspectral imagery which poses a heavy computational load due to its dense spectral bands. As data sizes continue to increase, computationally lightweight dimensionality reduction—such as offered by random projections—is likely to become ever more critical in hyperspectral applications.

In Chapter 2, we take the first stride in this direction by considering random projections based on a random Hadamard matrix (HM) as an alternative to the more widely used random Gaussian matrix (GM). HM-based projections ameliorate the computational burden by replacing a costly matrix multiplication with a series of addition and subtraction operations. We discuss in detail the similarities between the computationally efficient block-based-structure of the HM and that of the Discrete Fourier transform (DFT).

In Chapter 3, we further explore the realm of random projections by coupling them with more traditional transform-based methods in the form of a two-stage dimensionality reduction using a fast singular value decomposition (FSVD). More specifically, we first

employ a random projection to reduce dimensionality into an intermediate-dimensional space, and then perform FSVD-based feature selection to achieve reduction to the final low-dimensional space. Typically, transform-based dimensionality reduction methods are computationally intense but provide a desirable data-dependent dimensionality reduction that can capture relevant data structures. The two-stage dimensionality reduction thus leverages random projections in order to expedite the computation of the subsequent data-dependent FSVD transform. Ultimately, this yields an approximation to the true SVD at a fraction of the computational cost.

Finally, in Chapter 4, we address the commonly encountered spectral-unmixing problem by employing dimensionality reduction using random projection along with a nonnegative-least-squares (NNLS) process to yield a computationally efficient estimation of endmember abundances. The random-projection methods explored in the preceding chapters provided a computationally efficient reduction of dimensionality with NNLS being deployed in the resulting low-dimensional space. By several measures, the random-projection-based methods match or outperform NNLS deployed directly on the original dataset while running significantly faster.

Such is the overarching observation made generally throughout this dissertation—that random projections offer an efficient and easily implementable methodology for dimensionality reduction that permits hyperspectral-analysis tasks—such as unmixing and classification—to be conducted in a lower-dimensional space. In doing so, little, if any, performance is lost in the analysis task, while reductions in computational complexity

are significant. For this reason, random projections are anticipated to continue to be of paramount importance for the dimensionality reduction of hyperspectral imagery.

REFERENCES

[1] D. Achlioptas, "Database-Friendly Random Projections: Johnson-Lindenstrauss With Binary Coins," *Journal of Computer and System Science*, vol. 66, no. 4, June 2003, pp. 671–687.

[2] N. Ailon and B. Chazelle, "The Fast Johnson-Lindenstrauss Transform and Approximate Nearest Neighbors," *SIAM Journal on Computing*, vol. 39, no. 1, May 2009, pp. 302–322.

[3] N. Ailon and E. Liberty, "Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes," *Discrete & Computational Geometry*, vol. 42, no. 4, December 2009, pp. 615–630.

[4] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Authomatic Control*, vol. 19, no. 6, December 1974, pp. 716–723.

[5] R. Archibald and G. Fann, "Feature Selection and Classification of Hyperspectral Images With Support Vector Machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, October 2007, pp. 674–677.

[6] P. Bajcsy and P. Groves, "Methodology for Hyperspectral Band Selection," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 7, July 2004, pp. 793–802.

[7] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of Hyperspectral Images with Regularized Linear Discriminant Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, March 2009, pp. 862–873.

[8] A. Bannerjee, P. Burlina, and C. Diehl, "A Support Vector Method for Anomaly Detection in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, August 2006, pp. 2282–2291.

[9] R. G. Baraniuk, "Compressive Sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, July 2007, pp. 118–121.

[10] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach," *Neural Computation*, vol. 12, no. 10, October 2000, pp. 2385–2404.

[11] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, no. 6, June 2003, pp. 1373–1396.

[12] M. W. Berry, M. Browne, A. N. Langville, V. P. Paucac, and R. J. Plemmons, "Algorithms and Applications for Approximate Nonnegative Matrix Factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, September 2007, pp. 155–173.

[13] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral Subspace Identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, August 2008, pp. 2435–2445.

[14] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, April 2012, pp. 354–379.

[15] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping Target Signatures Via Partial Unmixing of AVIRIS Data," *Summaries of the Fifth Annual JPL Airborne Geoscience Workshop*, Pasadena, CA, 1995, pp. 23–26.

[16] C. Boutsidis and P. Drineas, "Random Projections for the Nonnegative Least-Squares Problem," *Linear Algebra and its Applications*, vol. 431, no. 5-7, August 2009, pp. 760–771.

[17] C. Boutsidis, A. Zouzias, and P. Drineas, "Random Projections for k-means Clustering," *Advances in Neural Information Processing Systems*, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds., Vancouver, Canada, December 2010, pp. 298–306.

[18] C. Boutsidis, A. Zouzias, M. W. Mahoney, and P. Drineas, "Randomized Dimensionality Reduction for k-means Clustering," *IEEE Transactions on Information Theory*, vol. 61, no. 2, February 2015, pp. 1045–1062.

[19] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, June 1998, pp. 121–167.

[20] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, June 2005, pp. 1351–1362.

[21] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite Kernels for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, January 2006, pp. 93–97.

[22] E. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, February 2006, pp. 489–509.

[23] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, March 2008, pp. 21–30.

[24] C.-I. Chang and D. C. Heinz, "Constrained Subpixel Target Detection for Remotely Sensed Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, May 2000, pp. 1144–1159.

[25] B. Chen, H. Liu, and Z. Bao, "A Kernel Optimization Method Based on the Localized Kernel Fisher Criterion," *Pattern Recognition*, vol. 41, no. 3, March 2008, pp. 1098–1109.

[26] C. Cortes and V. N. Vapnik, "Support Vector Networks," *Machine Learning*, vol. 20, no. 3, September 1995, pp. 273–297.

[27] S. Dasgupta, "Experiments with Random Projections," *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, June 2000, pp. 143–151.

[28] S. Dasgupta and A. Gupta, "An Elementary Proof of a Theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, January 2003, pp. 60–65.

[29] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal Processing With Compressive Measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, April 2010, pp. 445–460.

[30] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The Smashed Filter for Compressive Classification and Target Recognition," *Computational Imaging V*, San Jose, CA, January 2007, Proc. SPIE 6498, p. 64980H.

[31] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and Efficient Compressive Sensing Using Structurally Random Matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, January 2012, pp. 139–154.

[32] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian Endmember Extraction and Linear Unmixing for Hyperspectral Imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, November 2009, pp. 4355–4368.

[33] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, April 2006, pp. 1289–1306.

[34] Q. Du and J. E. Fowler, "Low-Complexity Principal Component Analysis for Hyperspectral Image Compression," *International Journal of High Performance Computing Applications*, vol. 22, no. 4, November 2008, pp. 438–448.

[35] Q. Du and J. E. Fowler, "On the Performance of Random-Projection-Based Dimensionality Reduction for Endmember Extraction," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Honolulu, HI, July 2010, pp. 1277–1280.

[36] Q. Du, J. E. Fowler, and B. Ma, "Random-Projection-based Dimensionality Reduction and Decision Fusion for Hyperspectral Target Detection," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Vancouver, Canada, July 2011, pp. 1790–1793.

[37] Q. Du and H. Yang, "Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, October 2008, pp. 564–568.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2 edition, John Wiley & Sons, Inc., New York, 2001.

[39] R. J. Durrant and A. Kaban, "Compressed Fisher Linear Discriminant Analysis: Classification of Randomly Projected Data," *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, Washington, DC, July 2010, pp. 1119–1128.

[40] M. D. Farrell and R. M. Mersereau, "On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, April 2005, pp. 192–195.

[41] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, November 2008, pp. 3804–3814.

[42] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision Fusion for the Classification of Urban Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, October 2006, pp. 2828–2838.

[43] X. Z. Fern and C. E. Brodley, "Random Projections for High Dimensional Data Clustering: A Clustering Ensemble Approach," *Proceedings of the International Conference on Machine Learning*, Washington, DC, August 2003, pp. 186–193.

[44] J. E. Fowler, "Compressive-Projection Principal Component Analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 10, October 2009, pp. 2230–2242.

[45] J. E. Fowler and Q. Du, "Reconstructions from Compressive Random Projections of Hyperspectral Imagery," *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, S. Prasad, L. M. Bruce, and J. Chanussot, eds., Springer, 2011, chapter 3, pp. 31–48.

[46] J. E. Fowler and Q. Du, "Anomaly Detection and Reconstruction from Random Projections," *IEEE Transactions on Image Processing*, vol. 21, no. 1, January 2012, pp. 184–195.

[47] J. E. Fowler, Q. Du, W. Zhu, and N. H. Younan, "Classification Performance of Random-Projection-Based Dimensionality Reduction of Hyperspectral Imagery," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Capetown, South Africa, July 2009, vol. 5, pp. 76–79.

[48] D. Fradkin and D. Madigan, "Experiments with Random Projection for Machine Learning," *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, Washington, DC, August 2003, pp. 517–522.

[49] P. Gamba, "A Collection of Data for Urban Area Characterization," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, September 2004, vol. 1, pp. 69–72.

[50] N. García-Pedrajas and D. Ortiz-Boyer, "Improving Multiclass Pattern Recognition by the Combination of Two Strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, June 2006, pp. 1001–1006.

[51] H. Goldberg, H. Kwon, and N. M. Nasrabadi, "Kernel Eigenspace Separation Transform for Subspace Anomaly Detection in Hyperspectral Imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, October 2007, pp. 581–585.

[52] Y. Gu, Y. Liu, and Y. Zhang, "A Selective KPCA Algorithm Based on High-Order Statistics for Anomaly Detecton in Hyperspectral Imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, January 2008, pp. 43–47.

[53] X. He and P. Niyogi, "Locality Preserving Projections," *Advances in Neural Information Processing System*, S. Thrun, L. Saul, and B. Schölkopf, eds., MIT Press, Cambridge, MA, 2004.

[54] D. C. Heinz and C.-I. Chang, "Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, March 2001, pp. 529–545.

[55] G. F. Hughes, "On the Mean Accuracy of Statistical Pattern Recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, January 1968, pp. 55–63.

[56] S. Jia and Y. Qian, "Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, January 2009, pp. 161–173.

[57] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz Mappings into a Hilbert Space," *Contemporary Mathematics*, vol. 26, 1984, pp. 189–206.

[58] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[59] V. C. Klema and A. J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, April 1980, pp. 164–176.

[60] K. Krishnamurthy, M. Raginsky, and R. Willett, "Hyperspectral Target Detection from Incoherent Projections: Nonequiprobable Targets and Inhomogeneous SNR," *Proceedings of the International Conference on Image Processing*, Hong Kong, September 2010, pp. 1357–1360.

[61] D. A. Landgrebe, "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, January 2002, pp. 17–28.

[62] C. Lee and D. A. Landgrebe, "Analyzing High-Dimensional Multispectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 4, July 1993, pp. 792–800.

[63] C. Li, T. Sun, K. F. Kelly, and Y. Zhang, "A Compressive Sensing and Unmixing Scheme for Hyperspectral Data Processing," *IEEE Transactions on Image Processing*, vol. 21, no. 3, March 2012, pp. 1200–1210.

[64] W. Li and J. E. Fowler, "Decoder-Side Dimensionality Determination for Compressive-Projection Principal Component Analysis of Hyperspectral Data," *Proceedings of the International Conference on Image Processing*, Brussels, Belgium, September 2011, pp. 329–332.

[65] W. Li, S. Prasad, and J. E. Fowler, "Classification and Reconstruction from Random Projections for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, February 2013, pp. 833–843.

[66] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Class Dependent Compressive-Projection Principal Component Analysis For Hyperspectral Image Reconstruction," *Proceedings of the Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Lisbon, Portugal, June 2011.

[67] J. Liu and J. Zhang, "Spectral Unmixing via Compressive Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, November 2014, pp. 7099–7110.

[68] X. Liu, W. Xia, B. Wang, and L. Zhang, "An Approach Based on Constrained Nonnegative Matrix Factorization to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, February 2011, pp. 757–772.

[69] N. H. Ly, Q. Du, and J. E. Fowler, "Reconstruction from Random Projections of Hyperspectral Imagery with Spectral and Spatial Partitioning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, April 2013, pp. 466–472.

[70] C. D. Martin and M. A. Porter, "The Extraordinary SVD," *American Mathematical Monthly*, vol. 119, no. 10, December 2012, pp. 838–851.

[71] G. Martín, J. M. Bioucas-Dias, and A. Plaza, "HYCA: A New Technique for Hyperspectral Compressive," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, May 2015, pp. 2819–2831.

[72] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, February 2001, pp. 228–233.

[73] S. Mei, Q. Du, and M. He, "Equivalent-Sparse Unmixing Through Spatial and Spectral Constrained Endmember Selection From an Image-Derived Spectral Library," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, June 2015, pp. 2665–2675.

[74] V. Menon, Q. Du, and J. E. Fowler, "Fast SVD with Random Hadamard Projection for Hyperspectral Dimensionality Reduction," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 9, September 2016, pp. 1275–1279.

[75] V. Menon, Q. Du, and J. E. Fowler, "Hadamard-Walsh Random Projection for Hyperspectral Image Classification," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Beijing, China, July 2016, pp. 5141–5144.

[76] V. Menon, Q. Du, and J. E. Fowler, "Random Projection Based Nonnegative Least Squares for Hyperspectral Image Unmixing," *Proceedings of the Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Los Angeles, CA, August 2016.

[77] V. Menon, S. Prasad, and J. E. Fowler, "Hyperspectral Classification Using a Composite Kernel Driven by Nearest-Neighbor Spatial Features," *Proceedings of the International Conference on Image Processing*, Québec City, Canada, September 2015, pp. 2100–2104.

[78] L. Miao and H. Qi, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, March 2007, pp. 765–777.

[79] J. M. P. Nascimento and J. M. Bioucas Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, April 2005, pp. 989–910.

[80] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Upper Saddle River, NJ, 1980.

[81] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas, "Random Projections for Support Vector Machines," *Proceedings of the* 16^{th} *International Conference on Artificial Intelligence and Statistics*, Scottsdale, AZ, April 2013, pp. 498–506.

[82] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas, "Random Projections for Linear Support Vector Machines," *ACM Transactions on Knowledge Discovery from Data*, vol. 8, no. 4, October 2014, Article No. 22.

[83] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent Advances in Techniques for Hyperspectral Image Processing," *Remote Sensing of Environment*, vol. 113, no. SUPPL. 1, September 2009, pp. S110–S122.

[84] S. Prasad and L. M. Bruce, "Limitations of Principal Component Analysis for Hyperspectral Target Recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, October 2008, pp. 625–629.

[85] K. I. Ranney and M. Soumekh, "Hyperspectral Anomaly Detection Within the Signal Subspace," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 3, July 2006, pp. 312–316.

[86] L. J. Rickard, R. W. Basedow, E. F. Zalewski, P. R. Silverglate, and M. Landers, "HYDICE: An Airborne System for Hyperspectral Imaging," *Imaging Spectrometry of the Terrestrial Environment*, G. Vane, ed., Orlando, FL, April 1993, Proc. SPIE 1937, pp. 173–179.

[87] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, December 2000, pp. 2323–2326.

[88] P. Sajda, S. Du, and L. Parra, "Recovery of Constituent Spectra Using Non-Negative Matrix Factorization," *Wavelets: Applications in Signal and Image Processing X*, M. A. Unser, A. Aldroubi, and A. F. Laine, eds., San Diego, CA, July 2003, Proc. SPIE 5207, pp. 321–331.

[89] T. Sarlós, "Improved Approximation Algorithms for Large Matrices via Random Projections," *Proceedings of the* 47^{th} *Annual IEEE Symposium on Foundations of Computer Science*, Berkeley, CA, October 2006, pp. 143–152.

[90] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, March 1978, pp. 461–464.

[91] J. Shang, R. Neville, K. Staenz, L. Sun, B. Morris, and P. Howarth, "Comparison of Fully Constrained and Weakly Constrained Unmixing Through Mine-Tailing Composition Mapping," *Canadian Journal of Remote Sensing*, vol. 34, no. sup1, 2008, pp. S92–S109.

[92] S. Subramanian, N. Gat, A. Ratcliff, and M. Eismann, "Real-Time Hyperspectral Data Compression Using Principal Components Transformation," *Proceedings of the AVIRIS Earth Science & Applications Workshop*, Pasadena, CA, February 2000.

[93] M. Sugiyama, "Dimensionaltity Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," *Journal of Machine Learning Research*, vol. 8, no. 5, May 2007, pp. 1027–1061.

[94] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple Spectral-Spatial Classification Approach for Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, November 2010, pp. 4122–4132.

[95] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[96] M. E. Winter, "N-FINDR: An Algorithm for Fast Autonomous Spectral End-Member Determination in Hyperspectral Data," *Image Spectrometry V*, M. R. Descour and S. S. Shen, eds. Proc. SPIE 3753, October 1999, pp. 266–275.

[97] A. Zare, P. Gader, and K. S. Gurumoorthy, "Directly Measuring Material Proportions Using Hyperspectral Compressive Sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, May 2012, pp. 323–327.